NEW FRAMEWORK FOR PROBABILISTIC INTERDEPENDENCY MODELING AND CRITICAL COMPONENT IDENTIFICATION TO INCREASE INFRASTRUCTURE SYSTEM RESILIENCE

A Dissertation
Presented to
The Academic Faculty

By

Chloe Johansen Applegate

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Civil Engineering

Georgia Institute of Technology

August, 2018

NEW FRAMEWORK FOR PROBABILISTIC INTERDEPENDENCY MODELING AND CRITICAL COMPONENT IDENTIFICATION TO INCREASE INFRASTRUCTURE SYSTEM RESILIENCE

Approved by:

Dr. Iris Tien, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. John Taylor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. John Crittenden
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Brani Vidakovic
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Leonardo Dueñas-Osorio
Department of Civil and Environmental Engineering
Rice University

Date Approved: July 25, 2018

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my advisor, Dr. Iris Tien, for her support, encouragement, mentorship, and advice throughout my time at Georgia Tech. Her expansive knowledge and drive have been critical to my success. I greatly appreciate the time she invested in me through weekly meetings, the I-Corps experience, and conferences. I am excited to have worked on research that I was passionate about under an extremely smart and kind mentor with Dr. Tien.

I would also like to extend gratitude to my thesis committee – Dr. John Crittenden, Dr. Leonardo Dueñas-Osorio, Dr. John Taylor, and Dr. Brani Vidakovic – for their insight and guidance on my thesis. I appreciate their contributions through coursework, data, and feedback on my thesis and research directions.

I would like to thank my fellow research group members – Cynthia Lee, Ajay Saini, Yanjie Tong, and Albert Zhang – for their friendship and support with my research. I appreciate the time each of you have put into helping find a way forward when my research felt stuck.

Finally, I am grateful for my family who have helped tremendously throughout my life. I extend gratitude to my husband, Spencer, parents, Greg and Dianne, and sister, Chelsae, who each gave me constant support, encouragement and love throughout my seemingly never-ending life as a student.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

acc_m	Access node m
BN	Bayesian network
c	Connection vector
$\mathbf{C}_{\boldsymbol{a_b}}$	components comprising MLS_a
ch	Unvisited children nodes of start component in **con**
C_i^B	Betweenness centrality for a node, i
C_i^C	Closeness centrality for a node, i
C_i^D	Degree centrality for a node, i
C_i^M	MLS appearances for a node, i
C_m	Component with access node as parent
$C_{m\,prev}$	Component with access node as parent's state in previous time step
con	Connectivity matrix of G
CPT	Conditional probability table
C_t	Transshipment or distribution component
$\mathbf{C}_{\mathbf{w}_x}$	Components comprising MLS_w
$\mathbf{C}_{\mathbf{Z}_j}^{\boldsymbol{k}}$	Components in zone Z_j
D_c	Comparative cutoff distance
d_{ij}	Length of shortest path between nodes i and j
D_S	Shortest distance from any supply node to the target node
G	Graph
H	Hazard node
L	Matrix of link lengths

L_C Current path length

LOS Percent level of service

$\mathbf{L_{rem}}$ Matrix of removed links

M MLS in the network

m Multiplier for maximum physical distance of MLS

MLS Minimum link set

MLS_a Cyclic MLS

MLS_{d_v} MLSs that are parents of component C_t

MLS_w Non-cyclic MLS

N Number of components in network

n_{a_b} Number of components in MLS_a

n_{con_i} Number of connections of node i

n_d Number of non-supply components

newVis New visited vector in recursive algorithm

n_H Number of hazard zones

n_{jk} Number of shortest paths between nodes j and k

$n_{jk(i)}$ Number of shortest paths between nodes j and k that contain the node i

n_M Number of MLS parent nodes of component C_t

n_{M_C} Number of cyclic MLSs

n_{MLS} Total number of MLSs in the network

$n_{M_{NC}}$ Number of non-cyclic MLSs

n_S Number of supply nodes

n_{w_x} Number of components in MLS_w

n_Z Number of zones

n_{Z_j}	Number of components in zone Z_j
$\mathbf{P_C}$	Current path vector
$p_{f_{b\|haz}}$	Probability of failure of components $\mathbf{C}_{a_b}$ given a hazard occurs
$p_{f_{b\|no\ haz}}$	Probability of failure of components $\mathbf{C}_{a_b}$ given a hazard does not occur
p_{f_m}	Probability of failure of component C_m
$p_{f_{q\|haz}}$	Probability of failure of component q given a hazard occurs
$p_{f_{q\|no\ haz}}$	Probability of failure of component q given a hazard does not occur
$p_{f_{t\|haz}}$	Probability of failure of component C_t given a hazard occurs
$p_{f_{t\|no\ haz}}$	Probability of failure of component C_t given a hazard does not occur
p_{H_i}	Probability of occurrence of hazard H_i
$p_{MLS\ cyc}$	joint probability of the calculated for MLS_a
p_{repair_m}	Probability of repair of component C_m
PMU	Power management unit
RAW	Risk achievement worth
RRW	Risk reduction worth
R_s	Service provision interdependency parent
S	Start node
$\mathbf{s}$	Vector of supply components
SCADA	Supervisory control and data acquisition
S_q	Supply component
T	Target node
$\mathbf{Vis}$	Visited vector
W_i^A	RAW on an interval scale for a node, i

$\widehat{W}_i^A$	RAW on a ratio scale for a node, i
W_i^{AD}	Combination RAW and degree metric for a node, i
W_i^{AM}	Combination RAW and MLS appearances for a node, i
w_i^D	Weight for node i based on degree
w_i^M	Weight for node i based on MLS appearances
W_i^R	RRW on an interval scale for a node, i
$\widehat{W}_i^R$	RRW on a ratio scale for a node, i
y	Number of removed links for a cyclic MLS_A
Z	Zone node

SUMMARY

Cascading failures of interdependent infrastructure networks have become increasingly critical as revealed by recent natural disasters and human disruptions. By determining how interdependencies both negatively and positively affect the fragility of infrastructure systems, it is proposed in this study that one can identify the most critical components and links, determine which infrastructure components to reinforce, and decrease the time required to regain normal operations of infrastructure systems.

With aging infrastructure and limited resources, a solution that systematically models the interdependencies between infrastructures can highlight areas that would most benefit from investment while accounting for the complex relationships between systems. There are sixteen critical infrastructure sectors defined by the U.S. Department of Homeland Security: chemicals, commercial facilities, communications, critical manufacturing, dams, defense, emergency services, energy, financial services, food and agriculture, government facilities, healthcare, information technology, nuclear facilities, transportation, and water and wastewater (White House, 2013). Each of these sectors are integrally tied to one another – many sectors depend on energy to power operation, the agriculture industry is tied to water infrastructure, and each sector must integrate with emergency services to prepare for potential emergencies. Interdependent power and drinking water infrastructure systems are analyzed in this study, and transportation and communication networks are discussed. These systems are analyzed because they are flow-based and can straightforwardly benefit from the methodology proposed.

Cascading failures, in which disruptions in one network lead to disruptions in connected networks, significantly increase the negative impacts of events such as blackouts, equipment failures, natural disasters, and organized attacks. For this reason, interdependencies are analyzed. Interdependencies are defined as relationships between two or more different infrastructure systems.

This dissertation presents a modeling approach and the accompanying sets of algorithms that enable computationally efficient probabilistic modeling of large infrastructure systems while considering interdependencies between networks. The proposed method creates a computationally tractable, representative Bayesian network of the system, with which exact inferences over single-component states in the network of interdependent systems are possible. Once the Bayesian network is constructed, inference analyses can be performed over a range of component state and hazard event scenarios to identify vulnerabilities across the network.

The model is applied to analyze component criticality within the infrastructure systems. Centrality-based and reliability-based component importance measures are considered. Centrality-based measures include degree centrality and those based on the system topology. Reliability-based measures include risk achievement worth, which define the impact of a single component outage on the probability of failure of the entire system.

The proposed methodology is applied to assess critical water services in the City of Atlanta, Georgia, including dependencies of the water distribution system on the power distribution network. Outcomes from a recent interdependent outage event and performance analyses can be used to validate the model. Repair, replacement, and reinforcement of infrastructure components can then be prioritize based on the model. New components may be revealed as critical when external dependencies are considered in the network of interdependent systems. Therefore, the importance of considering interdependencies in critical component identification is analyzed to understand how component criticality changes when making repair or investment decisions for infrastructure components when external dependencies are taken into account.

CHAPTER 1. INTRODUCTION

1.1 Motivation

Several factors have led to the necessity of addressing the resilience of America's interdependent infrastructure: (1) aging and antiquated infrastructure systems, (2) an increase in the risk of natural and anthropogenic hazards, and (3) an increase in effects of cascading failures. Critical infrastructure is defined as "a network of independent, mostly privately-owned, manmade systems and processes that function collaboratively and synergistically to produce and distribute a continuous flow of essential goods and services" (PCCIP, 1997). Several of these critical infrastructure systems are power, water, natural gas, communication, fuel, and transportation networks; each of which are inherently tied to each other through interdependencies. Interdependencies refer to relationships between two or more different infrastructure systems. Section 3.5 further defines interdependencies and the specific relationships between infrastructure systems that are investigated.

Infrastructure in the United States is aging. Many infrastructure components are at or near the end of their useful lives, including numerous drinking water pipes that were installed in the early to mid-20^{th} century and were expected to last up to 100 years. An estimated 240,000 water main breaks occur annually and between 14-18% of water is estimated to be lost daily due to leaky pipes. Some parts of the electric grid were built before the turn of the 20^{th} century and many transmission and distribution lines were constructed in the mid-20^{th} century with expected useful lives of 50 years (American Society of Civil Engineers, 2017). Financial restraints and a lack of available components render the magnitude of necessary repairs infeasible. For example, in 2010, the power transformer manufacturers in the United States met approximately 15% of the demand for power transformers in the country. The lead time between ordering a transformer and receiving it ranged from five to twelve months (U.S. Department of Energy, 2014) as many of the transformer components are imported from China and Europe. In an emergency situation, it is not feasible to get multiple power transformers to a location to quickly bring facilities back online.

With population growth and rapid urbanization, vulnerability and exposure to disasters is increasing because more people and assets are in areas of high risk (UNIDSR, WMO, 2012). More people are living in areas with high earthquake and flood risk. Additionally, cyber adversaries are advancing in capability and are creating an increased threat to information systems that all infrastructure systems rely on. Cyber-attacks, such as SYNful Knock in 2015, have allowed cyber adversaries access to infrastructure operator credentials, and therefore access to control infrastructure devices. This can allow adversaries to control the network infrastructure, redirect traffic, and cause damage to and large outages in infrastructure systems (United States - Computer Emergency Readiness Team, 2016). These phenomena reveal the necessity to focus on resilience – "the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions" (White House, 2013) – through preparation before a hazard occurs, effective response during a hazard, and rapid recovery after a hazard occurs.

Catastrophic consequences of events such as hurricanes Katrina in 2005, Sandy in 2012, and Harvey in 2017 and the Haiti and Tohuku earthquakes in 2010 and 2011, respectively, have demonstrated the impact that these hazards can have on infrastructure systems. Hurricane Sandy, for example, caused power outages for close to two million people, damage to nearly 800 buildings, a gasoline shortage, closure of six hospitals, internet, cable, and cell phone outages for up to 11 days, flooding of all six subway tunnels connecting Brooklyn and Manhattan, interruption to commutes of 217,000 vehicles, over 500 miles of damaged roads, and release of partially treated or untreated sewage from 10 of 14 wastewater treatment plants in New York alone (PlaNYC, 2013). Many of these outages were due to or exacerbated by the power outages that occurred, displaying a cascading failure throughout many infrastructure sectors. Another example of a cascading failure is the 2003 Northeast blackout, the cause of which was a tree coming in contact with power lines, which led to overloads and outages throughout the power grid in the Northeastern United States and Ontario, Canada (Andersson et al., 2003). The cascading effects were disruptions of the drinking water, transportation, and communication networks because of their dependence on the power network (Hernandez-Fajardo & Dueñas-Osorio, 2013). To mitigate this problem for future hazards, system reinforcement and recovery for resilience must be conducted with a focus on the connections and interdependencies that exist between infrastructure systems.

The methodology and framework proposed in this study have been developed to address each of these concerns. The methodology can be used to help infrastructure system owners and decision makers prioritize repairs and replacements of aging components, understand potential impacts of hazards to reinforce systems and decrease effects of those hazards, and identify potential points of cascading failures to prevent their occurrence. All of this can be done in a computationally efficient manner without making approximating assumptions about the states of components. The model accounts for all connections and relationships between infrastructure components when performing inference. The model is made using a Bayesian network (BN), which accounts for uncertainty and allows for component-level updates to propagate across the network.

1.2 Outline

The remainder of this dissertation is as follows:

Chapter 2 provides the background for the proposed methods for interdependent infrastructure system modeling and analysis. The chapter starts with an analysis of existing methods for modeling interdependent infrastructure systems. Then the definition and applications of BNs are discussed, followed by existing component importance measures. Finally, resilience is introduced.

Chapter 3 introduces the novel methodology proposed to create a model of interdependent infrastructure systems. Each step is discussed thoroughly, including the model inputs, dimensionality reduction to allow computationally tractable inferences, definition of interdependencies to account for complex relationships between infrastructure systems, construction of the adjacency matrix to define the BN structure, discussion of cycles that arise in the formulation of the BN, and definition of conditional probability tables that define the probabilities of components being in each possible state.

Chapter 4 evaluates the application of the methodology to an example water distribution system and its dependencies on power distribution in Atlanta, Georgia. Each step of the methodology proposed Chapter 3 is described in the application. The time taken to run several steps of the methodology are also discussed.

Chapter 5 presents the application of the methodology to perform inference and identify critical components in the network. The different applications are presented, followed by a discussion of the component importance measures that are used in the application described in Chapter 4. Two component importance measures combining two previously defined metrics are proposed. Finally, the component rankings are discussed along with their application to increasing infrastructure system resilience.

Chapter 6 provides an evaluation of the performance of the proposed methodology. A comparison is performed to an existing method to model infrastructure systems and a comparison of the results is provided.

Chapter 7 summarizes the main contributions of this work. Potential further applications of this work are discussed.

CHAPTER 2. BACKGROUND

2.1 Introduction

This chapter provides the background for the proposed methodology for modeling infrastructure networks and the interdependencies between them. The chapter contains a review of methods to model interdependent infrastructure networks, and a definition of BNs and their applications. This is followed by a discussion of existing component importance measures and their applications to infrastructure systems. Finally, resilience is defined and its application to the proposed model is discussed.

2.2 Interdependent Infrastructure Networks

Interdependencies refer to relationships between two or more different infrastructure systems, such as between power and natural gas or water and power. Natural gas supplies generation capacity to power plants and power supplies the electricity needed to operate water treatment plants and pump stations.

2.2.1 Existing Methods to Define Interdependencies

One study proposing interdependency types was performed by Rinaldi, Peerenboom, and Kelly (2001). The authors define four types of interdependencies: physical, cyber, geographic, and logical. Physical interdependency refers to a physical linkage between the inputs and outputs of two infrastructure systems. An example of a physical interdependency refers to a rail network providing coal for fuel to a coal-fired power plant, while the rail network relies on electricity from the power plant to operate. Cyber interdependency is defined when the state of a component in one infrastructure system is determined based on information transmitted to communication infrastructure. Modern infrastructure systems rely on supervisory control and data acquisition (SCADA) systems and power management units (PMUs) for management and operations, which create a cyber interdependency. Geographic interdependencies arise when a local hazard causes similar changes to components in close proximity. An example of geographic interdependency is the failure of several components in an area due to an explosion or fire. Logical interdependencies refer to all other connections between infrastructure systems that are not physical, cyber, or geographic.

2.2.1.1 Limitations of Existing Methods to Define Interdependencies

To specifically address infrastructure resilience, explicit and comprehensive definitions are necessary that relate to all aspects of resilience. The logical interdependency defined by Rinaldi, Peerenboom, and Kelly (2001) is unclear and would be covered by one of the interdependency types defined in Section 3.5 of this dissertation. Additionally, the interdependency types defined by Rinaldi, Peerenboom, and Kelly (2001) do not relate to the post-hazard recovery aspect of infrastructure resilience. In this study, a new type of interdependency is proposed – access for repair – which is an important addition for a focus on resilient infrastructure systems. Access for repair interdependency is defined in Section 3.5.3.

2.2.2 Existing Methods to Model Interdependent Infrastructure Networks

There are many approaches to modeling interdependencies between critical infrastructure systems. These include empirical, agent-based, system dynamics-based, economic theory-based, and network-based approaches (Ouyang, 2014). The method proposed in this study is a network-based approach – one where nodes represent different infrastructure components and links represent the connections between them. Network-based approaches are able to analyze system components considering all capacities of resilience – resistance, absorption, and restoration. Adaptive capacities are also considered. Resistance is the ability for infrastructure systems to prevent and withstand potential hazards, prior to the hazard occurring. Absorption refers to lessening the effects of a hazard during the event, including taking actions to accelerate decision making in the case of an emergency and utilizing system redundancies. Restoration refers to activities to support recovery, including community notifications and optimized sequences of response. Adaptive capacities include increasing the strength of infrastructures and installing monitoring for the states of systems to decrease vulnerability to future disasters. Network-based approaches are effective at evaluating the ability of the network to prevent events that lead to large consequences, determining the effects of improving absorptive capacities of critical infrastructure components, and analyzing how well the network supports advanced design decisions to quickly find restoration priorities (Ouyang, 2014).

Previous work in identifying and accounting for interdependencies in infrastructure networks include Chou & Tseng (2010) and Halfawy (2008). In Chou & Tseng (2010), failure records of different infrastructure types are used to predict interdependencies through sequence-based failure events. Halfawy (2008) focuses on how to integrate management of multiple municipalities to optimize asset management decisions over multiple infrastructure types that may have different owners. Another approach to analyzing infrastructure interdependencies is the inoperability input-output model. This model analyzes how disruptions to one infrastructure system propagate to other infrastructure systems through the exchange of input and output resources that are transferred between systems (Satumtira & Dueñas-Osorio, 2010).

2.2.2.1 Limitations of Existing Methods to Model Interdependent Infrastructure Networks

Methods to identify and model infrastructure interdependencies, such as those proposed by Chou & Tseng (2010) and Halfawy (2008) can be easily integrated into the proposed framework. New interdependencies learned or predicted can be added to the model through the defined interdependency relationships to assess potential cascading failures. The results from the models proposed can be used across infrastructure owners to address priorities in investment to mutually benefit multiple infrastructure stakeholders. Inoperability input-output models are typically applied to account for economic interdependencies between infrastructure systems (Akhtar & Santos, 2013; Santos, et al., 2014). If desired, nodes representing economic variables can be added to the proposed framework, both at the component and system levels.

While network-based approaches enable identification, description, and analysis of most resilience strategies, they can require a large quantity of data input to generate the network graph. Prior applications of BNs have typically focused on only one or a few systems. In addition, the BN model's ability for updating – new information entered at any node in the BN propagates to all nodes in the network – addresses the limitations of other static approaches, such as input-output-based methods proposed by Leontief (1951) and

Rose and Miernyk (1989). Static approaches describe the state of a system at one point in time rather than allowing for updating as components age and change.

2.3 Bayesian Networks

2.3.1 Overview of BNs

BNs model systems to account for the probabilistic dependencies between components and facilitate updating of system assessments with new information. The BN is a directed (i.e., edges are directional) and acyclic (i.e., no closed path exists in the network) probabilistic graph composed of nodes and links. Based on the dependency relationships between components, nodes are defined as parent and children nodes. Children depend on the states of their parents. Each node represents a random variable and for discrete networks, where components have categorical states, is defined by a conditional probability table (CPT). For variables with parent nodes, the CPT consists of the conditional probabilities of the states of the child node given the states of the parents. For variables without parent nodes, the CPT consists of the marginal probabilities. A BN can be updated, meaning that new information entered at any node in the BN propagates to all nodes in the network.

Specifically, the BN framework allows for incorporation of both prior and updating information. Prior knowledge about each component is added to the BN during construction of the network. When new information is learned about a component, including through measurements and observations, it is updated, with the effects propagated to all other nodes in the system through inference. For example, if a failing component is replaced, a decrease in that component's probability of failure is likely to decrease the probability of failure of connected components.

2.3.2 Existing Applications of BNs

2.3.2.1 BNs in Computing Applications

In computing applications in civil engineering, BNs have been used in several ways. One is to identify damage location on civil structures using E/M impedance (Naidu, Soh, & Pagalthivarthi, 2006). In that paper, BNs are used to reduce the amount of input data needed for traditional damage identification methods, which require large amounts of training data. Cheng & Hoang (2014) probabilistically estimate slope stability using BNs to calculate posterior probabilities of slope collapse without requiring prior knowledge of data distributions. As in these studies, BNs are useful in the approach described in this dissertation because a large amount of input data is not necessary to learn information about the network and to calculate probabilities of failure of infrastructure components based on different scenarios.

2.3.2.2 BNs for Single Infrastructure Networks

BNs have been used to model single infrastructure networks such as inland waterway ports (Hosseini & Barker, 2016), railway lines (Castillo, Grande, & Calviño, 2016), highways (Grande, Castillo, Mora, & Lo, 2017), power (Tien & Der Kiureghian, 2017) and water networks (Leu & Bui, 2016). These studies have not considered interdependencies between different networks. In Leu & Bui (2016), the BN nodes are defined based on general properties of the water network (e.g., pipe diameter and depth) and other factors that could affect the water network (e.g., pipe corrosion and construction activities). Hosseini & Barker (2016) build a BN model where resilience metrics, such as backup utility systems and quick evacuation, are nodes in the network. A BN was used to analyze the risk of domino effects, similar to cascading failures, in chemical plant infrastructure in Khakzad (2015). Nodes represent parts of a fuel storage plant, such as tanks, which can be in the states of safe, on fire, or burned out. The cascading failures modeled are in time slices, applying a dynamic BN model. Similarly, a dynamic BN was used to evaluate cascading effects in the power grid in Codetta-Raiteri, et al. (2012). In that study, electrical lines are considered as series or parallel modules that connect nodes in the power grid.

In general, past studies used BNs to analyze small infrastructure systems of five to ten components (e.g., Bobbio et al., 2001; Kim, 2011); the algorithms proposed in this dissertation allow BNs to be used for much larger systems of hundreds of components. Another application of BNs for infrastructure reliability assessment is in Bensi, Der Kiureghian, and Straub (2013), where an efficient modeling algorithm was developed to create chain-like BN structures to model infrastructure systems. The BN is modeled using survival and failure path chain-like events, where the state of the event depends on the states of preceding events in the chain. Mahadevan, Zhang, and Smith (2001) applied BNs to assess the reliability of structural systems accounting for multiple failure sequences and correlations between component-level limit states.

2.3.2.3 BNs for Interdependent Critical Infrastructure

BNs have also been used to model the security of interdependent critical infrastructure (Schaberreiter, Bouvry, Röning, & Khadraoui, 2013). This approach uses service outputs and high-level system measurements as nodes in the network. Aung & Watanabe (2010) similarly model interdependent infrastructure systems using BNs at a very high level, where each node in the network represents an entire critical infrastructure sector. The BN is used to determine cascading effects of infrastructure sector outages. The critical infrastructure BN model in Di Giorgio & Liberati (2011) also includes nodes representing services supplied and single nodes representing infrastructure systems, e.g., the electrical transmission system as a whole, along with nodes representing adverse events.

2.3.3 Limitations of Existing Applications of BNs

Compared to previous studies using BNs to model infrastructure systems, the focus of this study is on large, complex infrastructure networks, accounting for the states of individual components of each system. In prior studies where BNs were used to model single infrastructure networks, nodes in the BN were used to represent properties of a network and other global factors that affect the network. The proposed approach instead uses nodes to represent the states of the individual components in the network and links to represent the connectivity between them. In infrastructure networks, overall system states are governed by individual component states. The methodology in this study enables consideration of the states of specific components, whose performance impacts overall infrastructure system performance. The resulting model can then be used to analyze diverse scenarios, including component-level events, with levels of service outcomes measuring resilience of the network under different conditions. Because of computational demands, previous BN-based single network approaches have not considered interdependencies between the networks they are modeling and other networks on which they depend.

Previous studies modeling security of interdependent infrastructure systems using BNs differ from the framework proposed in that rather than modeling, as single nodes, entire infrastructure systems or just the services provided, the proposed approach models infrastructures starting from the constituent components of a system. This dissertation considers both the level of the individual infrastructure components and the topology and connectivity characteristics of infrastructure networks. In practice, this is where the complex relationships between systems, including interdependencies between them, arise. For example, the probability of being able to provide a service at a distribution component is dependent on the number and reliabilities of redundant paths, which are themselves composed of other components, to that distribution point. For a water distribution system, that probability is also dependent on the reliability of the power components supplying electricity for the water treatment plants and pump stations.

In contrast with previous studies, each node in the proposed BN model represents an individual component of an infrastructure network. This enables a user to capture component-to-component relationships, as well as incorporate any component-level information, such as updates about component states from monitoring or inspection information, into system assessments. In addition, decisions for infrastructure systems occur at the component level, e.g., which component should have an additional backup, or between which components should a redundant path be built. The proposed framework supports these component-level inferences. Resulting analyses allow infrastructure owners to identify specific nodes, representing individual components, in the network considered critical for replacement, repair, or additional buildouts to increase overall system performance.

The methodology proposed uses a BN-based approach to capture probabilistic relationships between components and incorporate both prior information about the network and update assessments when new information is learned about the network. Prior information is incorporated during construction of the BN. Updating information is incorporated during inference of the BN. For example, if it is learned that a certain hazard occurs on the system or a certain component fails, the new information is propagated to all nodes in the network to calculate updated probabilities across all component and system states.

2.4 Component Importance

In reliability engineering, component importance measures are used to identify areas for prioritized investment in infrastructure networks (Baroud & Barker, 2018). Previous work has highlighted the importance of considering interdependencies in vulnerability analyses of critical infrastructure systems. Johansson & Hassel (2010) point to functional and geographical interdependencies as particularly important. In this study, service provision, geographic, and access for repair interdependencies are defined to comprehensively capture the connections between infrastructures. Component importance measures are explored to prioritize investment in interdependent networks rather than single infrastructure systems as in Baroud & Barker (2018). Both functional interdependencies, defined as service provision, and geographical interdependencies are considered here. In this way, the impacts of considering realistic infrastructure interdependencies are accounted for when making decisions on the components to repair, replace, or reinforce to improve overall network reliability. For example, components may have low criticality when considering only one infrastructure system, but can have cascading effects when considering the interdependent systems.

What follows is an overview of two broad classes of component importance measures – centrality-based and reliability-based metrics – and evaluation of their use for critical component identification in infrastructure systems. The use is considered for specific infrastructure applications, the background for their selection in this work, and limitations of prior studies in the area. In general, centrality metrics consider the layout of an infrastructure network, but do not account for reliability of individual components. Reliability-based measures include component reliability, but less explicitly include network topology.

2.4.1 Centrality-Based Metrics

Several deterministic approaches to centrality-based metrics exist, and the following metrics which have been used in analysis of infrastructure networks are investigated: betweenness centrality, closeness centrality, degree centrality and the number of appearances in minimum link sets (MLSs). These metrics are chosen based on the literature review and their applicability to particular infrastructure sectors. Degree centrality relates to the immediate risk of a node being influenced by the state of the resource flowing through the network. For example, a node in a water network with many connections, and therefore a high degree, has higher risk of being affected by a contaminant in the water network. Previously, minimum cut sets have been used to represent system reliability in single networks (Espiritu, Coit, & Prakash, 2007). Here, MLS appearances are used to represent component criticality in interdependent networks as they are widely applicable to any flow-based network. The number of appearances in MLSs indicates the number of critical paths for which a component is a part to deliver a resource from supply to distribution nodes.

Other metrics, such as betweenness centrality and closeness centrality are also commonly applied. Betweenness centrality represents how many shortest paths throughout the network contain a node. Closeness centrality represents the speed by which information can be exchanged throughout a network (Comfort & Haase, 2006). Table 1 summarizes how each metric is calculated (expanded upon in the corresponding sections), the meaning of the metric, and the infrastructure sectors to which the suggested metric is to be applied. The metrics are further explained, including definitions of variables and rationale for application, in sections 2.4.1.1 to 2.4.1.4.

Table 1. Summary of centrality-based metrics.

Metric	Calculation	Definition	Sectors
Betweenness Centrality	$C_i^B = \sum_{j,k \in G, j \neq k \neq i} \frac{n_{jk(i)}}{n_{jk}}$ (1)	• Proportion of shortest paths in network that contain node i • Nodes with high betweenness have more control in the network	• Communication • Power
Closeness Centrality	$C_i^C = \frac{N-1}{\sum_{j \in G; j \neq i} d_{ij}}$ (2)	• Sum of length of all shortest paths in network • Closer nodes have better access to information or more direct influence on other vertices	• Communication
Degree Centrality	$C_i^D = \sum_{j=1}^{n} con_{ij}$ (3)	• Number of nodes connected to node i • Immediate risk of a node catching what is flowing through network	• Natural Gas • Transportation • Water
MLS Appearances	$C_i^M = \sum_{j=1}^{n_{MLS}} M_j$ (4)	• Number of MLSs that node i appears in • High number of MLSs indicates node is critical to supplying resource flowing though the network to many nodes	• Communications • Natural Gas • Power • Transportation • Water

2.4.1.1 Betweenness Centrality

Betweenness centrality (C_i^B) for a node, i, is defined as the proportion of times the shortest paths connecting two nodes in a network include node i. Betweenness centrality is calculated using equation (1),

$$C_i^B = \sum_{j,k \in G, j \neq k \neq i} \frac{n_{jk(i)}}{n_{jk}} \tag{1}$$

where G represents all of the nodes in the graph, n_{jk} represents the number of shortest paths between nodes j, and k and $n_{jk(i)}$ represents the number of shortest paths between nodes j and k that contain the node i (Crucitti, Latora, & Porta, 2006). Nodes with high betweenness centrality tend to contribute to multiple risk paths in the case of cascading failures (Stergiopoulos, Kotzanikolaou, Theocharidou, & Gritzalis, 2015).

Betweenness centrality is typically applied to roadways, railways, and power transmission networks. For example, in the power grid, betweenness is a proxy for the power transmitted through a substation (Albert, Albert, & Nakarado, 2004). Since nodes with high betweenness centrality indicate a higher level of control in the network, it is suggested that this metric is applied to interdependent networks communication and power systems. When assessing the relevance of different component importance measures for critical infrastructure systems, several assumptions made when using betweenness centrality make the metrics more appropriate for transportation and power networks compared to, e.g., water or gas distribution systems. The first assumption is that the resource flowing through the network is assumed to follow one of the shortest paths and the second is that, if there are several shortest paths, the probability of choosing each of those paths is assumed to be equal (Grubesic, Matisziw, Murray, & Snediker, 2008). For water and gas networks, other dynamics must be considered such as hydraulics and pressure zones throughout the system.

2.4.1.2 Closeness Centrality

Closeness centrality is a measure of how near a node is to all other nodes along shortest paths in a network (Crucitti, Latora, & Porta, 2006). Closeness centrality (C_i^C) for node i is calculated using equation (2):

$$C_i^C = \frac{N-1}{\sum_{j \in G; j \neq i} d_{ij}} \tag{2}$$

where N is the number of components in the graph (G) and d_{ij} is the length of the shortest path between nodes i and j.

Closeness centrality is used as a resilience measure in Omer, Mostashari, and Lindemann (2014), where a transportation system is analyzed for resilience and node criticality. Closeness centrality is used because it relates to the accessibility of a node compared to the rest of the network and resilience is thought of as the ability to access nodes that have failed in a hazard. Closeness centrality can also represent the speed by which information can be exchanged throughout a network (Comfort & Haase, 2006). For this reason, closeness centrality is applicable to networks such as communication where speed is necessary in response activities.

2.4.1.3 Degree Centrality

Degree centrality of node i (C_i^D) represents the number of edges that are connected to a node (Stergiopoulos, Kotzanikolaou, Theocharidou, & Gritzalis, 2015). Equation (3) is used to calculate the degree centrality of node i,

$$C_i^D = \sum_{j=1}^{N} \mathbf{con}_{ij} \tag{3}$$

where N is the number of nodes in the graph and $\mathbf{con}$ represents the connectivity matrix of the network. The entries in $\mathbf{con}$ are equal to one if nodes are connected and zero otherwise (Grubesic, Matisziw, Murray, & Snediker, 2008).

Nodes with high degree centrality represent nodes with high criticality. These are nodes that have a higher direct association with other nodes and can indicate that a node is a hub in a network (Grubesic, Matisziw, Murray, & Snediker, 2008). As will be later shown, this is consistent with the findings in this study, where nodes with a high degree are generally supply components or components that distribute resources to many other nodes in the network.

Degree centrality is often treated as one of the most important metrics for evaluating the robustness of a network. However, it is not as indicative of component criticality when there is a high level of parity in a network. For example, in Grubesic, Matisziw, Murray, and Snediker (2008), an internet network is analyzed and the degree centrality ranges from 2 to 3 for the entire network. This shows very little difference in importance, but other metrics show that some nodes are more critical than others. Therefore, degree centrality alone is not sufficient in assessing component importance.

2.4.1.4 Minimum Link Set Appearances

The number of MLSs in which a node appears (C_i^M) is viewed as another centrality metric. A MLS is a minimum set of components that must survive for a pathway to survive between two nodes in a network. MLSs for a given node represent the paths that must be functional in order for the node to obtain the resource flowing through the network. The number of MLS appearances as a centrality metric for node i is calculated using equation (4):

$$C_i^M = \sum_{j=1}^{n_{MLS}} M_j$$
$$\text{where } M_j = \begin{cases} 1 \text{ if } M_j \text{ contains component } i \\ 0 \text{ otherwise} \end{cases} \tag{4}$$

where M_j represents each MLS in the network and n_{MLS} denotes the total number of MLSs in the network, obtained from the MLS formulation, discussed in Section 3.4.1.

It is intuitive that a node that appears in many MLSs is more critical than a node appearing in few MLSs since the node is necessary to carry the resource flowing through the network from a supply point to a transshipment or distribution node.

2.4.1.5 Existing Applications of Centrality-Based Metrics

Dunn and Wilkinson (2013) apply graph-theoretic measures, such as betweenness and degree centrality, to a hydraulic model of a water network and find that a combination of physically-based and graph-theoretic component importance measures better predicts component criticality than physically-based measures alone. Compared to this work, this study assesses the impacts of interdependencies on component importance rather than attempting to identify critical components within a single infrastructure network alone. Physically-based measures account for either the hydraulic flow through the network or the physical relationships between components. Graph-theoretic measures relate to centrality-based metrics as discussed in this dissertation. Combination metrics are proposed that account for both physically-based measures – risk achievement worth in this case – and graph theoretic measures – centrality metrics.

Stergiopoulos et al. (2015) analyze betweenness centrality, eccentricity, closeness centrality, Eigenvector centrality, and degree centrality for their application to risk analysis in interdependent infrastructure systems. The study analyzes dependency risk paths and calculates the information gain for each of these metrics. Degree centrality and closeness centrality are determined to be the most accurate indicators of nodes with high impacts on risk in a network. Closeness centrality represents the speed by which information can be exchanged throughout a network (Comfort & Haase, 2006). For this reason, closeness centrality is applicable to networks such as communication and transportation where speed is necessary in response activities. Stergiopoulos et al. (2015) do not consider individual component reliabilities in their assessments of criticality, however. However, as individual component performance is critical in contributing to overall network reliability, these values are considered in calculations of importance.

Zhang, Miller-Hooks, and Denny (2015) apply graph-theoretic network measures on 17 common network structures for transportation systems. The network measures include connectivity measures, such as cyclicity and degree, and accessibility measures, such as diameter and betweenness centrality. The common network structures refer to the topology of the transportation network and include grid networks, hub and spoke, and crossing paths. Cyclicity, average degree, and diameter are used to analyze the resilience of transportation networks. The application of degree is analyzed in this study and recommend its use for transportation networks.

Several of these studies have applied centrality-based component importance measures to single infrastructure networks without consideration of external dependencies (Pinnaka, Yarlgadda, and Çetinkaya, 2015; Zhang, Miller-Hooks, and Denny, 2015). However, they provide a basis for the recommendations made for the use of specific metrics in certain infrastructure sectors. Stergiopoulos et al. (2015) considers interdependencies in their use of centrality-based component importance measures. However, the outcome of the study is to indicate critical sectors rather than critical components within each network.

2.4.2 Reliability-Based Metrics

In contrast to centrality-based measures, reliability-based metrics take into account varying reliabilities of individual components in the network. Vesely and Davis (1985) discuss two measures of importance to be used for probabilistic risk analyses – risk reduction worth (RRW) and risk achievement worth (RAW). These measures can be used to prioritize quality and risk assurance programs for plant management and regulation (Vesely & Davis, 1985). These metrics are applied as the canonical reliability-based measures to assess component criticality. They are calculated using BNs, where the evidence of each component's outage is input and the change in system failure probability can be observed. Espiritu, Coit, and Prakash (2007) apply two complementary metrics to RRW and RAW – reliability reduction worth and reliability achievement worth – to commonly used electrical configurations in power networks. The metrics are transformed for electrical transmission systems to include outage rates for components of the transmission system. Reliability reduction worth and reliability achievement worth provide a lower and upper limit on component importance, respectively. Because RAW is both the lower limit of reliability when ranking component importance and has less parity when components have low probability of failure, RAW is suggested for use with all infrastructure sectors. Table 2 provides a summary of the reliability-based metrics discussed – RRW and RAW. These is expanded upon in sections 2.4.2.1 and 2.4.2.2.

Table 2. Summary of reliability-based metrics.

Metric	Calculation	Definition	Sectors
Risk Reduction Worth	$\widehat{W}_i^R = \frac{\text{P(System Failure \| Survival of Node } i)}{\text{P(System Failure)}}$ (5) $W_i^R = P(\text{System Failure \| Survival of Node } i) - P(\text{System Failure})$ (6)	• Maximum decrease in risk with an improvement to node i	• Communications • Power • Natural gas • Transportation • Water
Risk Achievement Worth	$\widehat{W}_i^A = \frac{\text{P(System Failure)}}{\text{P(System Failure \| Failure of Node } i)}$ (7) $W_i^A = P(\text{System Failure}) - P(\text{System Failure \| Failure of Node } i)$ (8)	• Maximum increase in risk with failure of node i	• Communications • Power • Natural gas • Transportation • Water

2.4.2.1 Risk Reduction Worth

Risk reduction worth quantifies how much a particular node can reduce the risk level of a network. RRW can be calculated on a ratio scale or an interval scale. RRW on ratio scale ($\widehat{W}_i^R$) is shown in equation (5) and RRW on interval scale (W_i^R) is shown in equation (6):

$$\widehat{W}_i^R = \frac{\text{P(System Failure | Survival of Node } i)}{\text{P(System Failure)}} \quad (5)$$

$$W_i^R = \mathrm{P(System\ Failure\ |\ Survival\ of\ Node\ } i) - \mathrm{P(System\ Failure)} \tag{6}$$

where $\mathrm{P(System\ Failure\ |\ Survival\ of\ Node\ } i)$ is the conditional probability of failure of the infrastructure system given that node i has 100% probability of survival and $\mathrm{P(System\ Failure)}$ is the original probability of failure of the infrastructure system (Vesely & Davis, 1985).

RRW can be used to prioritize component improvements to reduce risk to the infrastructure system (Oliveira, Mota De Sá, & Ferreira, 2014). However, RRW can have a high degree of parity when components in a network have very low probabilities of failure. When this occurs, there is little distinction between the criticality of different components.

2.4.2.2 Risk Achievement Worth

Risk achievement worth measures the overall increase in risk at the system level if a component node has failed. RAW for a node i can be calculated on ratio scale ($\widehat{W}_i^A$) and on interval scale (W_i^A), shown in equations (7) and (8), respectively:

$$\widehat{W}_i^A = \frac{\mathrm{P(System\ Failure)}}{\mathrm{P(System\ Failure\ |\ Failure\ of\ Node\ } i)} \tag{7}$$

$$W_i^A = \mathrm{P(System\ Failure)} - \mathrm{P(System\ Failure\ |\ Failure\ of\ Node\ } i) \tag{8}$$

where $\mathrm{P(System\ Failure)}$ is the prior probability of system failure $\mathrm{P(System\ Failure} \mid \mathrm{Failure\ of\ Node\ } i)$ is the conditional probability of failure of the infrastructure system given that node i has 100% probability of failure (Vesely & Davis, 1985).

In this study, RAW is calculated on the interval scale. The prior probabilities of failure of each component are calculated using the built probabilistic model of the infrastructure network. Then, inference is performed by updating the network with 100% probability of failure for component node i. Posterior probabilities of failure for all components other than i are calculated. Finally, the differences between the prior and posterior probabilities of failure for all components are summed.

RAW is related to maintenance impact on a component because it quantifies the worth of restoring a failed component (Vesely & Davis, 1985). RAW can be used to prioritize the most important components for reliability assurance and maintenance decisions (Oliveira, Mota De Sá, & Ferreira, 2014, August). As previously described, RAW provides the lower limit of reliability when ranking component importance and displays less parity for systems with low component failure probabilities. Therefore, it is suggested that RAW is applied to all infrastructure sectors.

2.4.2.3 Existing Applications of Reliability-Based Metrics

Reliability-based metrics, including RRW and RAW, have previously been applied to single infrastructure systems including power (Volkanovski, Čepin, & Mavko, 2009), inland waterways (Baroud, Barker, & Ramirez-Marquez, 2013), and a generalized infrastructure network (Barker, Ramirez-Marquez, & Rocco, 2013). Multiple infrastructures are considered in Patterson and Apostolakis (2007), in which RRW and RAW are used to identify critical locations across multiple infrastructure systems. Rather than identifying critical components in the infrastructures, geographic regions are identified as critical. This does not account for the service provision interdependencies between multiple infrastructure systems.

2.5 Impact of Interdependencies

The individual component importance metrics above have previously been applied to single infrastructure networks. Many interdependencies exist within critical infrastructure systems, however, which can cause disruptions in key industries, lead to billions of dollars in lost productivity, and cause widespread security and reliability concerns. Infrastructure dependency connections can cause failures that are not captured when modeling infrastructures separately. For example, a water supply component such as a pump station depends on power in order to operate. When a power network is considered independently of the water system, the power distribution component that supplies the pump station may not be considered critical. However, when the water systems are considered, the power distribution component's criticality increases because it can cause a cascading failure in the water system. Therefore, this dissertation outlines the investigation of the impact of considering external dependencies in identifying critical components in infrastructure systems to make and prioritize decisions such as for maintenance and retrofit.

2.6 Resilience

Providing a quantitative way to measure the impact of critical infrastructure on communities is essential for improving resilience. Some literature regarding the quantification of seismic resilience for communities and infrastructure networks can be found in Bruneau et al. (2003), where measures of increased resilience are reduced failure probabilities and reduced consequences from failure; Franchin and Cavalieri (2015), in which resilience is based on displaced population, road damage, and recovery strategy; and Guidotti et al. (2016), where resilience is quantified based on the loss of functionality and delay in recovery.

Baroud, Barker, & Ramirez-Marquez (2014) also discuss costs and interdependent impacts of infrastructure network resilience. They have developed a network resilience metric that allows decision makers to assess resilience-building decisions after a disruption occurs. They optimize the time to network restoration, the time to restore full network service, and the total time spent on recovery activities.

This dissertation proposes methods to model and assess the performance of interdependent infrastructure systems under varying scenarios. This is accomplished through a probabilistic Bayesian network model, where inference gives updated probabilities of failure. This is useful to increase system resilience before a hazard occurs to assess where the greatest extent of damage is possible and where to invest resources to prevent large outages. The model is useful during a hazard to determine where to disperse resources and repair crews to bring the most customers or the most critical customers back online as quickly as possible. Finally, the model is useful after a hazard to prioritize components for interventions to prevent similar incidents and impacts from occurring again in the future.

This dissertation also proposes to rank component criticality for interdependent infrastructure networks. These rankings can be used to improve the resilience of infrastructure systems before, during, and after hazards occur. Before hazards, the framework can be used to identify vulnerabilities of the system in day-to-day infrastructure operations. During hazards, outages can be input into the model to identify potential cascading failures and to dispatch repair workers and resources to areas of the highest vulnerability or most critical assets. After a hazard, the model can be used to help determine a prioritized order of repair to get the most customers or assets back on line as quickly as possible.

CHAPTER 3. BAYESIAN NETWORK MODEL OF INTERDEPENDENT INFRASTRUCTURE NETWORKS

3.1 Introduction

This chapter provides a detailed description of the proposed framework to construct a BN model of interdependent infrastructure networks, including how the BN is constructed with required inputs, component connection models, and interdependency definitions within the framework. The approach and the accompanying novel algorithms are described, which both address challenges associated with BN modeling of complex infrastructure systems. These include algorithms for minimum link set identification in infrastructure systems, super-component identification, and modeling different types of interdependencies. These also include a method to identify and remove cyclic component configurations within the network. The overall methodology creates a computationally tractable probabilistic BN model of large infrastructure systems considering interdependencies between networks. The BN model can be used to calculate component importance measures in the network, and two metrics combining centrality- and reliability-based metrics are proposed in section 3.9.

3.2 Overview

Figure 1 shows a diagram of the overall methodology. To create the interdependent infrastructure system model, the methodology begins with inputs of component locations, connectivity, types, and failure probabilities. A binary system is considered in this case, where components can be in one of two possible states, e.g., failure or survival. Therefore, failure probabilities are defined. The method easily extends to systems of multiple states such as those modeling flow capacity. In those cases, probability distributions over possible components states would be defined (Tong & Tien, 2017).

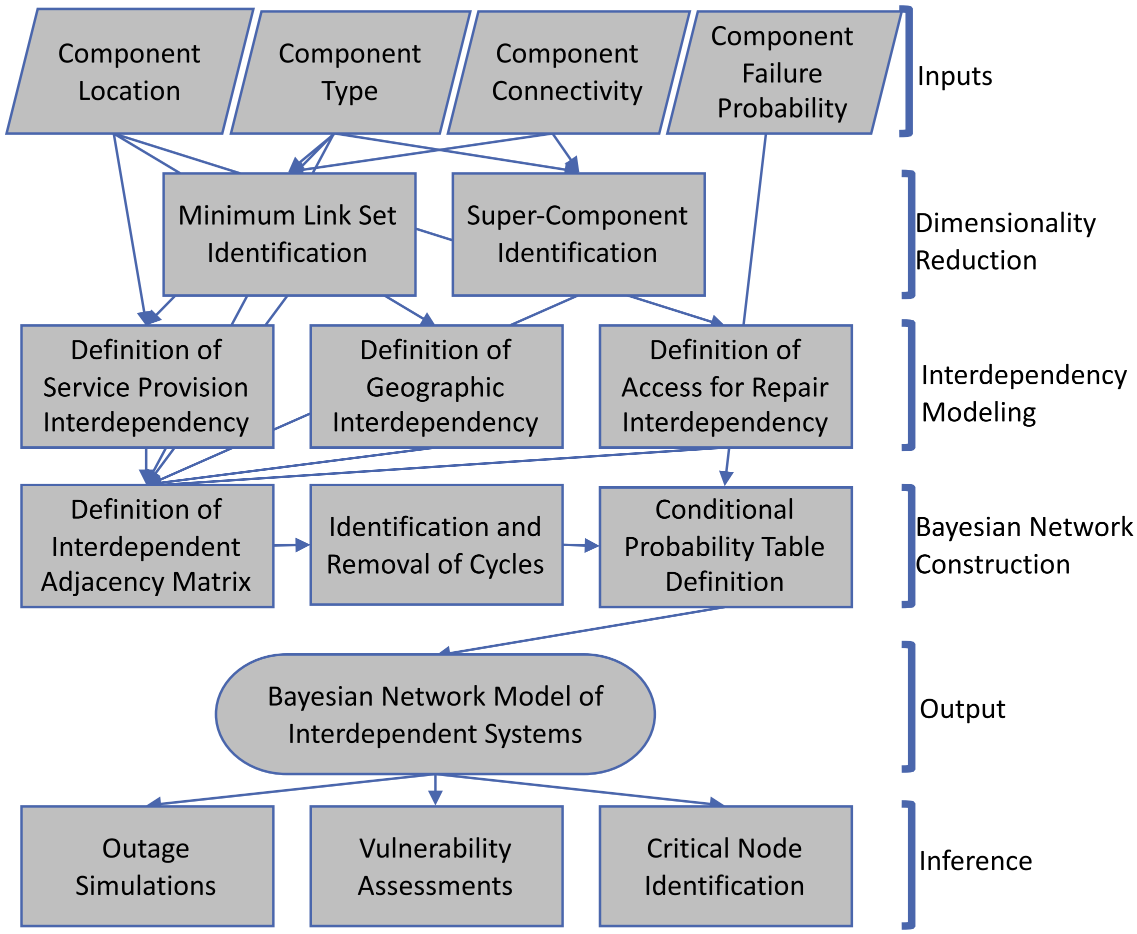


Figure 1. Diagram of overall methodology.

The next step is to reduce the dimensionality of the BN by using MLS formulation combined with super-component identification. Algorithms to do this in the context of infrastructure systems are proposed in section 3.4. Next, the interdependencies are defined and modeled, and the BN is constructed, including definition of the interdependent adjacency matrix, identification and removal cycles in the interdependent infrastructure network, and definition of conditional probability tables. The resulting model is then used to perform exact inference to probabilistically assess the vulnerability of interdependent infrastructure networks.

3.3 Inputs

The construction of the BN is largely based on geospatial information about the interdependent infrastructure networks. It is assumed that any given infrastructure system is composed of individual components whose performance contributes to the performance of the overall network. The required inputs for the methodology are component locations, connectivity, types, and failure probabilities.

3.3.1 Component Locations

The first inputs necessary are the locations of all components in the networks. For example, for a water distribution network, the locations include coordinates of supply components – water treatment plants, pump stations, tanks, and reservoirs – pipe junctions, and terminal distribution nodes. The locations provide information on the components that provide infrastructure resources or services to specific parts of the community. In addition, the methodology accounts for the relation between components and hazards that are considered for risk assessments, as most hazards can be described geospatially. For example, the distance from an earthquake fault line can be computed based on the component locations. Any specification of locations is acceptable as long as the frame of reference is the same across the networks and hazards.

3.3.2 Component Connectivity

The component connectivity inputs are expressed as pairs of infrastructure components that are connected in an individual network. For example, for a water distribution system, the connectivity is provided as a list of the two system components that each pipe connects. The connectivity inputs are translated into a connectivity matrix with each row and column representing a component. The matrix is composed of ones and zeros where a value of one signifies that the component in the row is connected to the component in the column and a zero signifies no connection.

3.3.3 Component Type

Another input needed is the type of each component. There are three component types considered in the methodology, each corresponding with a different physical function within an infrastructure system. The first is supply components. These are components that generate or output the resource that flows through the network. For example, for a water distribution system, the resource is water, and the supply nodes are considered water treatment plants, pump stations, tanks, and reservoirs. The second type is distribution components. These are endpoints in the network that distribute the infrastructure resource to customers or end users. These can also be smaller distribution points such as small pumps or power lines for the water and power systems, respectively, which feed to individual houses or facilities. The final type is transshipment components. These are intersections between several links in the system that are not endpoints in the network. These enable the infrastructure resource to be distributed more easily across the network along multiple paths compared to having single flow paths. The component type inputs are necessary to define the role of each component within overall system functioning.

3.3.4 Component Probability of Failure

The final input is failure probabilities for each component. These can be determined from empirical calculations, asset health scores, or estimated probabilities from infrastructure owners or other domain-specific experts. An example of empirical calculations is to use fragility curves to calculate failure probabilities given a specific hazard (González, Dueñas-Osorio, Sánchez-Silva, & Medaglia, 2016). In other cases, infrastructure owners calculate failure probabilities for each component, such as those corresponding with varying asset health scores.

If precise failure probability values cannot be determined, relative values can be used, particularly if a goal of the analysis is to provide a comparative ranking and prioritization of the system components. E.g., a uniform probability of failure can be assigned to all nodes in the network and the output allows a comparison across all components. These failure probabilities can also be updated when new information is learned about the system. For example, if the results of an inspection update the estimated probability distribution of the state of the component, that information updates the prior failure probability of the component, and through its connections, the distributions of the states of surrounding nodes as well.

3.4 Dimensionality Reduction

With these inputs, the dimensionality of the BN is then reduced using both a MLS formulation and super-component identification. These approaches make it computationally tractable to model systems of hundreds of nodes.

3.4.1 Minimum Link Set Formulation

In assessing the performance of a system, a MLS is a minimum set of components that are required to be functioning for the system to function. For a physical infrastructure system, a MLS includes the components that must be working for a resource, e.g. water, power, or gas, to be conveyed from a source node to any other node in the network. If a single component in the MLS fails, the MLS fails. For this application, MLSs link supply components to transshipment or distribution components. The MLS formulation used represents the connectivity of the network but is not based on flow or capacity of the network. They map paths through the network to provide infrastructure services to end points in communities. This can be done manually for small networks; however, this is time consuming for networks of larger than 10 components.

Complementary to MLSs are minimum cut sets (MCSs). For infrastructure networks, a MCS is the minimum set of components that must fail for a resource to fail to be conveyed from a source node to any other node in the network. Several algorithms have been developed to identify MCSs in networks. One of these algorithms is EG-CUT developed by Shin & Koh (1998). This builds a MCS generation tree and backtracks from a leaf when it fails to generate a MCS. However, this method does not enumerate MLSs.

A robust, efficient algorithm to define the MLSs of the system enables the capturing of the functionality of the network while reducing the dimensional complexity of the BN. It models the influence of every combination of individual component states on overall system performance through the MLSs. Previously, a recursive decomposition algorithm was proposed to identify shortest paths in a network (Li & He, 2002). However, this method does not scale with the size of the network (Lim, Song, & Kurtz, 2015). Here, a recursive algorithm based on a depth-first search method (Jiang, Bai, Atkin, & Kendall, 2017) is proposed in this study. For large infrastructure networks, a cutoff for the maximum size of a MLS can be incorporated based on the logic that a resource will not deviate far from the shortest path to increase computational efficiency. For example, water will not weave through a grid in a network to travel between two points on a single line. In applications where circuitous routes are forced (e.g., if valves throughout a cater system are closed), these cutoffs can be removed to account for each possible path.

The recursive MLS identification algorithm presented as **Algorithm A** is run for each pair of supply components and target components, which include all transshipment and distribution components in the network. Inputs to the algorithm are the start component (S), target component (T), connectivity matrix ($\mathbf{con}$), the shortest distance (D_S) from any supply component to the target component of interest, and a matrix of the physical length of all links in the network (**L**). Unbolded italics denote scalar values; small bolded letters denote vectors; capital bolded variables denote matrices. A comparative distance (D_C) is calculated to create a distance cutoff for the maximum physical distance of the MLS using a multiplier (m). In the application example in this dissertation, the cutoff distance is twice the shortest distance between the supply and target nodes, i.e., $m = 2$.

As the algorithm proceeds, it "visits" an increasing number of components. Several variables are created during the recursion of the algorithm. These include a visited vector $(\mathbf{Vis})_{1\times\mathrm{n}}$ of ones and zeros the length of the number of components (N) that represents the components that have been visited – one representing an unvisited component and zero representing a visited node during the course of the algorithm, a current path ($\mathbf{P_C}$) vector that represents the path calculated within the recursive algorithm, and a current length variable (L_C) that represents the length of the current path.

Algorithm A. MLS identification algorithm.

```
Input: S, T, Con, D_S, L, m, Vis, P_C, L_C
  D_c = m · D_S
  Vis(S) = 0
        P_C = [P_C, S]
  Add length of link to length of current path L_C
  ch = unvisited connections in con
  If S = T
     MLS = P_C
  Else for each element (i) in ch:
     S = ch(i)
     If L_C > D_C, break
     newVis = Vis
     newVis(S) = 0
     newPaths = MLSalg(S, T, con, newVis, P_C, L_C)
```

Variables are defined as:

- S = supply node,
- T = target node,
- **con** = connectivity matrix for the network,
- D_S = shortest distance from any supply node to the target node,
- **L** = matrix of link lengths,
- m = multiplier for maximum physical distance of MLS,
- **Vis** = $1\times$ n vector of visited nodes, initiated as all 1s representing that all nodes are unvisited,
- $\mathbf{P_C}$ = vector of components in the current path of the MLS, initiated as an empty vector representing that no nodes are yet included in the path,

- L_C = length of current path, initiated as 0 indicating that the current path length is zero, and
- $\mathbf{ch}$ = vector of children nodes of S found in $\mathbf{Con}$ not yet visited.

In the algorithm, first, a comparative distance (D_c) is calculated as a multiplier (m) times the shortest distance (D_S). Next, the start component (S) is marked as visited in the visited vector ($\mathbf{Vis}$). The start component is then added to the current path ($\mathbf{P_C}$). Unless the current path only contains one node, the length of the link added to the current path is added to the current length (L_C). The children ($\mathbf{ch}$) variable is defined as the children of the start component (S) – found in $\mathbf{con}$ – that have not been visited as found in $\mathbf{Vis}$. A MLS is discovered if the source component is equal to the target component ($S = T$) and is defined as the current path ($\mathbf{P_C}$). If an MLS is not found, the algorithm then cycles through each element (i) of the child vector ($\mathbf{ch}$) and sets the child as the source component (S). If the current path length (L_C) is greater than the cutoff distance (D_C), the algorithm moves onto the next supply node. Otherwise, new variables are defined to move onto the next recursion. The new visited vector is set ($\mathbf{newVis}$) and the new start component is marked as visited. Finally, the algorithm calls itself to repeat with the next S component until a MLS is reached or the algorithm has visited all elements of the $\mathbf{ch}$ vector.

3.4.2 Super-Components

A second method to reduce the dimensionality of the network is in using super-components (Der Kiureghian & Song, 2008; Tien, 2017), which combine multiple components to model them using a representative single node. One way to define a super-component is as a subset of components in the system that are connected in series or parallel (Tien & Der Kiureghian, 2017). These reduce the effective number of nodes in the BN while still representing the state of each node in the system (Bensi, Der Kiureghian, & Straub, 2013). Components that comprise the super-component are represented as parents of the super-component in the BN, reducing computational complexity of the model without making any approximating assumptions.

For the proposed algorithm, a super-component is defined when its state is known given the failure of any one of its constituent components. For this approach, components in a series configuration are grouped into a super-component. Those components become parents of a super-component node in the BN. The algorithm presented as **Algorithm B** is used to define the super-components using just the connectivity matrix ($\mathbf{con}$) and a vector of the supply components ($\mathbf{s}$) as inputs. The algorithm is as follows:

Algorithm B. Super-component identification algorithm.

```
Input: con, s
For each transshipment and distribution component in the network (i)
   c = con(i, :)
   If s ∉ c and length of c = 2
      For each element of c (j)
         If the length of con(j, :) = 2 , super-component identified
```

The algorithm loops over each non-supply component in the network (i) and defines the connection vector ($\mathbf{c}$) as all components indicated in the connectivity matrix with one values. Supply components are not included as their functionality differs from that of transshipment and distribution components. The algorithm then selects connections in $\mathbf{c}$ that have exactly two connections (length of $\mathbf{c} = 2$), representing a component in series. Finally, the algorithm loops through all elements of the connection vector ($\mathbf{c}$) and makes the same selection process to either add components to the current super-component or to identify a new super-component.

3.5 Defining Interdependencies

In the modeling methodology, three comprehensive interdependency types are included – service provision, geographic, and access for repair – that affect the resilience of infrastructure systems. These three interdependency types are used because they encompass the possible connections between infrastructures through a physical connection, connection by geography, and resilience-aspects of interdependence; and are well defined compared to previous definitions of interdependencies where a logical interdependency represents relationships not covered by other types.

3.5.1 Service Provision Interdependency

3.5.1.1 Definition of Service Provision Interdependency

In the proposed framework, a service provision interdependency refers to the function of a component in one system relying on the function of a component from another system. This covers both the physical and cyber interdependencies as defined by Rinaldi, Peerenboom, and Kelly (2001). An example of this is a water pump station requiring electricity from a power substation to function. If there is an outage at the power substation, the water pump station would fail, particularly if no backup power is present.

3.5.1.2 Modeling of Service Provision Interdependency

To model a service provision interdependency, a direct link is added from the supplying – parent – component providing the service to the dependent – child – component. For example, to model the dependency of a water pump station on power, a link is added from the power substation to the water pump station. An assumption made in building the BN is that the closest supplying component provides the resource to the dependent component (Dueñas-Osorio, Craig, & Goodno, 2007). The ability of the supplying component to provide the resource to the dependent component is represented in the BN. If other information on how resources are supplied across infrastructure networks is available, that information is easily incorporated into the BN through direct links between those components. Child components can have multiple parents. For example, water treatment plants often have feeds from multiple power substations. In that case, each of the substations is represented as a parent node. Components typically requiring a service provision interdependency include natural gas and water supply components that depend on power. Other components in each of these systems do not require power in order to function. For example, gravity-fed water distribution components, particularly in older systems, can usually operate without power. As systems become increasing automated, however, these service provision interdependencies will increase.

An example BN for a service provision interdependency is shown in Figure 2, where the service provision interdependencies are shown as dashed arrows. The example BN comprises a power system of components $\mathrm{C}_{1\mathrm{p}}$, …, $\mathrm{C}_{m\mathrm{p}}$, where m is the number of power components, and a water system of components $C_{1w}, \ldots, C_{nw}$, where n is the number of water components. The MLSs are numbered $MLS_{1p}, \ldots, MLS_{sp}$ and $MLS_{1w}, \ldots, MLS_{tw}$, where s represents the number of MLSs for the power network and t represents the number of MLSs for the water network. The service provisions interdependencies are from C_{mp} to C_{2w} and from C_{1w} to $C_{(m-1)p}$. The relationship from C_{mp} to C_{2w} can represent the dependence of a water pump station on a power distribution station, as in the above example. The relationship from C_{1w} to $C_{(m-1)p}$ can represent a power plant depending on water from a water distribution station.

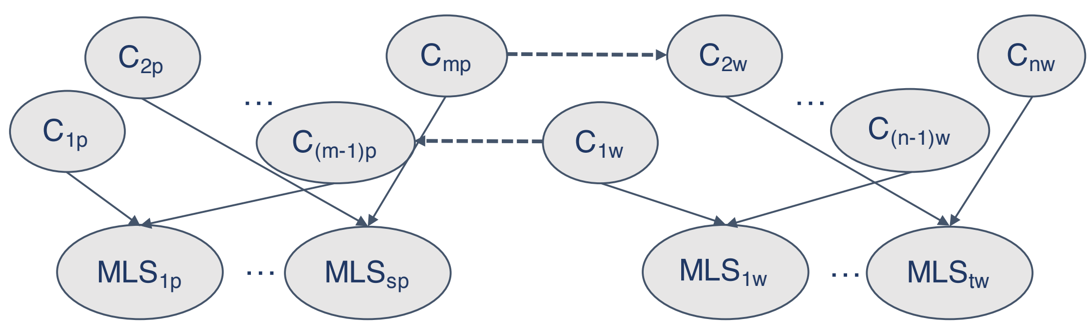


Figure 2. Example BN for service provision interdependency.

3.5.2 Geographic Interdependency

3.5.2.1 Definition of Geographic Interdependency

A geographic interdependency refers to the relationship between two or more components in the same geographic area that are likely to experience similar effects given a local hazard. This interdependency type is consistent with the geographic interdependency defined by Rinaldi, Peerenboom, and Kelly (2001). For example, components in proximity are likely to fail concurrently if a hazard were to occur in that area. It is common for gas and water lines to be routed along the same road so that only one trench is necessary; this represents a geographic interdependency where the two lines are more likely to fail together given a common hazard. Collocated components are also likely to be repaired together during the restoration process.

3.5.2.2 Modeling of Geographic Interdependency

To model a geographic interdependency, components are grouped into hazard zones. For each zone, a common hazard node is added, which is a parent of all of the components in the zone. Hazard nodes can be created for any set of components to capture multiple hazard impacts. These include cyber nodes that represent cyber threats on infrastructure systems, and natural disaster nodes representing earthquake or hurricane threats. Zones can be determined based on proximity to certain hazards, collocation of components, or service areas around supply components. An example using proximity to a hazard would be to partition components based on their distance from an earthquake fault line. An example to partition the components using service areas is to use a k-nearest neighbor search to group components by their closest supply nodes.

Figure 3 shows an example BN for a geographic interdependency, where the dashed lines represent the geographic interdependencies. Hazard zones are represented by $\mathrm{Hazard}_1$to Hazard_s, where s denotes the number of hazard zones. Components with the same hazard parents are in the same hazard zone. In the example, components $\mathrm{C}_{1\mathrm{p}}$, $\mathrm{C}_{1\mathrm{w}}$, $\mathrm{C}_{(\mathrm{m}-1)\mathrm{p}}$, and $\mathrm{C}_{(\mathrm{n}-1)\mathrm{w}}$ are all in hazard zone 1.

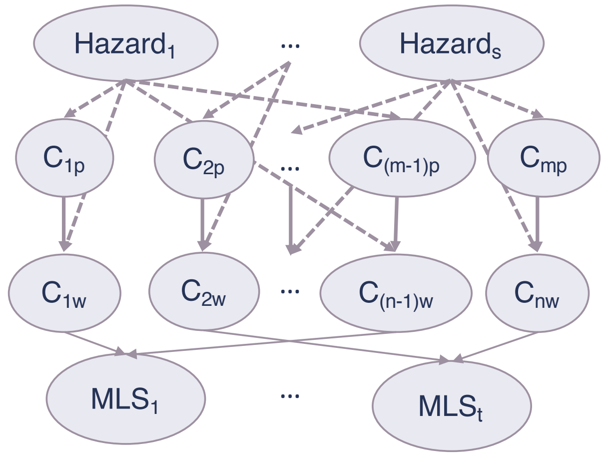


Figure 3. Example BN for geographic interdependency.

System nodes can also be created based on the geographic partitions that represent the resources provided to particular service areas when performing inference on the system. Each system node is a child of all of the components within that system, such as those within a geographic partition.

3.5.3 Access for Repair

3.5.3.1 Definition of Access for Repair Interdependency

An access for repair interdependency is defined for certain infrastructures that must be functioning to gain cyber or physical access to a failed component to repair it. For example, if a water network component loses function, communication systems will be necessary to report the failure or gather information about the event through monitoring systems. Transportation systems must also be working for repair crews to access the failed component. This interdependency type is proposed specifically to address infrastructure resilience, taking into account systems necessary for post-disaster recovery and restoration.

Rather than the undefined logical interdependency type listed by Rinaldi, Peerenboom, and Kelly (2001), three explicit interdependency types of service provision, geographic, and access for repair are applied in this methodology, which represent the possible connections that exist between infrastructure systems.

3.5.3.2 Modeling of Access for Repair Interdependency

When modeling an access for repair interdependency, the change in the operational status of infrastructure components over time is taken into account. Access nodes are created as parent nodes of the components that depend on them. Access nodes only affect the state of a child component in the case of component failure. Generally, a working component is independent, for example, of the state of its connected communication or transportation networks. Communication systems are used to provide information on the functioning of components to operators, however, this will not affect a component's survival or failure. Therefore, a node representing the state of the component in its previous time step is created, allowing the determination of the need to account for the state of an access node. In defining the access nodes, for cyber access, these nodes account for required communication with the dependent component and the robustness of the communication channels to disruptions. For physical access, the access nodes represent remoteness and redundancy in transportation paths to reach the component. The probability of repair of a component given access can be defined based on the criticality of component or the availability of resources for repair. Probability of repair may evolve over time. For example, if everything is operating normally, the probability of failure may be 100% as the infrastructure decision makers may have all of the resources needed to repair a failed component on hand. However, if a hazard has occurred, probabilities of repair will vary based on the criticality and accessibility of components. In this case, the probability of repair can be updated and input into the BN. Inference using the proposed methodology takes approximately four seconds on a network with 119 components.

An example BN is shown in Figure 4 with C_{1w} as the potentially failed node and the access nodes defined as a telecommunications tower and road providing cyber and physical access, respectively. $C_{1w\ previous}$ represents the state of the water node in the previous time step.

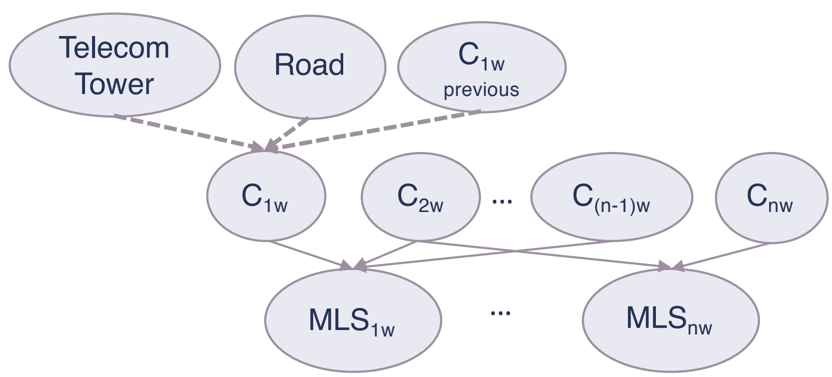


Figure 4. Example BN for access for repair interdependency.

3.6 Interdependent Adjacency Matrix Construction

The structure of a BN is defined by an adjacency matrix. This is similar to a connectivity matrix where it is composed of ones and zeros. Here, a value of one indicates the connection between a parent node in the row and a child node in the column and zero indicates no connection. In addition, rather than just capturing the original network topology with components as in the connectivity matrix, the adjacency matrix includes the defined MLSs, super-components, and interdependency relationships. The adjacency matrix also accounts for directionality of the dependency and is no longer symmetric. The adjacency matrix is constructed based on the parent-child relationships defined in the previous steps. For MLSs in particular, each component that comprises an MLS is a parent of the MLS node, and each MLS node is a parent of the component for which it is an MLS to provide the resource to that component. This functional relationship introduces potential cycles in the graph. The method to address these cycles is presented in the following section.

3.7 Accounting for Cycles

In the modeling and assessment of complex infrastructure systems, it is possible for cycles to arise in the creation of the BN graph. For example, suppose component C1 in the water network, denoted $\mathrm{C_{1w}}$, is a part of a MLS for component $\mathrm{C_{nw}}$. Suppose at the same time that, based on the topology of the network, $\mathrm{C_{nw}}$ is a part of a MLS for component $\mathrm{C_{1w}}$. With these two components each a part of the other's MLSs, a cycle is introduced. BNs, however, must be acyclic graphs. Typically, the system would no longer be able to be modeled as a BN. Here, a novel algorithm has been developed as a method to identify and remove the cycles in the BN, discussed in the following, and account for the dependency that the cycle introduces, discussed in Section 3.8.5.2.

An example of a cycle that arises from the MLS formulation is shown in Figure 5, where component $\mathrm{C_1}$ is part of the MLS ($\mathrm{MLS_2}$) for $\mathrm{C_2}$ and $\mathrm{C_2}$ is included in the MLS ($\mathrm{MLS_1}$) for $\mathrm{C_1}$. A method to remove the cycles in the graph while retaining the dependency relationships between the nodes is presented. Specifically, the components are defined by their joint probability distribution and with a removal of any one of the links from a component to a MLS within the cycle, thus removing the cycle. In the example, the link from C_1 to MLS_2 is removed, shown with the dotted line. The joint probabilities are calculated by considering the possible configurations of each component state and using total probability to calculate the remaining values. If all components comprising MLS_2 are functional, the state of MLS_2 then depends on the state of C_1. The probability of failure of C_1 is then calculated using the probabilities of failure of its parents. This calculation is further described in **Algorithm C** presented later in this section to define the MLS conditional probability tables.

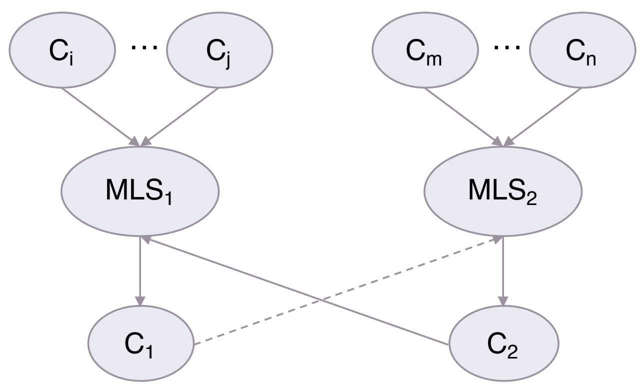


Figure 5. Example BN with cyclic dependency.

In this step, all cycles in the graph must be identified. This is performed by traversing the graph represented by the interdependent adjacency matrix and identifying when the path traversed reaches a previously visited node. This represents a cycle in the graph. Once the cycle is identified, a link from a component to a MLS node is removed in the adjacency matrix.

When traversing the graph, the process begins with component C_i. A visited vector is constructed as $[C_i]$. The graph traverse algorithm then moves to C_i's connections, in this case MLS_1. The visited vector is now $[C_i, MLS_1]$. This continues, traversing to each connection until the visited vector reaches $[C_i, MLS_1, C_1, MLS_2, C_2, MLS_1]$. In this step, a node in the visited vector is repeated, indicating a cycle. At this step, a link is removed in the adjacency matrix, in this case the link from C_1 to MLS_2, by setting that entry to zero. The removed link is noted to define the conditional probability distribution in the next step.

3.8 Defining Conditional Probability Tables

Each node in the BN must be defined by a conditional probability distribution of its state given the states of its parents, typically represented for discrete or discretized variables in a CPT. The calculation of the CPT varies for each node type. Details for the CPT calculation for nodes defining geographic interdependencies, access for repair interdependencies, supply components including service provision interdependencies, transshipment and distribution components, and MLSs both non-cyclic and cyclic are provided in this section.

3.8.1 Geographic Interdependency Nodes

When defining the geographic interdependency in the network, hazard nodes are created that represent components in the same physical area that are likely to experience correlated outcomes in the case of a hazard. Hazard nodes do not have parent nodes, so their CPTs are defined as the marginal probabilities of hazard occurrence, as shown in equation (9), where H_i, $i = 1, \ldots, n_H$, represents a hazard node and n_H is the number of hazard zones. The probability of occurrence of hazard H_i is p_{H_i}.

$$\begin{gathered} \mathrm{P}(H_i \text{ occurs}) = p_{H_i} \\ \mathrm{P}(H_i \text{ does not occur}) = 1 - p_{H_i} \end{gathered} \tag{9}$$

To assess overall performance of the infrastructure systems, zone nodes can be incorporated to account for levels of service provided in each service area in the community. Zone nodes represent these service areas, with the parents of zone nodes including all of the nodes within the zone. Equation (10) defines one example of a calculation of CPTs for zone nodes, where Z_j, $j = 1, \ldots, n_Z$, represents a zone node and n_Z is the number of zones. The components within each zone are represented by $\mathbf{C}_{\mathbf{Z}_j}^{\boldsymbol{k}}$, $k = 1, \ldots, n_{Z_j}$ where n_{Z_j} is the number of components in zone Z_j. In equation (10), the percent level of service is denoted as LOS. A zone's level of service is defined as the percentage of components that are in the working state within the zone's partition. A zone node can also be defined by the proportion of its customers that have continued service or by the reliability of service in a zone.

$$\mathrm{P}\big(Z_j \text{ is at } LOS\% \text{ Service}\big) = \begin{cases} 1 \text{ if } LOS\% \text{ of } \mathbf{C}_{\mathbf{Z}_j}^{\boldsymbol{k}}, k = 1, \ldots, n_{Z_j} \text{ survive} \\ \quad 0 \text{ otherwise} \end{cases} \tag{10}$$

3.8.2 Access for Repair Interdependency Nodes

Access nodes are created when defining access for repair interdependencies. This is performed for all supply, transshipment, and distribution components in the networks that depend on them for repair. Equation (11) defines a component CPT when it has an access node as a parent. The component is denoted C_m, the access node acc_m, and the component's state in the previous time step $C_{m\,prev}$. The probability of failure of component C_m is p_{f_m}. The probability of repairing component C_m is defined as p_{repair_m}, which is described in Section 3.5.3.2.

$$P(C_m \text{ survives}) = \begin{cases} 1 - p_{f_m} \text{ if } acc_m \text{ survives and } C_{m\,prev} \text{ survives} \\ 1 - p_{f_m} \text{ if } acc_m \text{ fails and } C_{m\,prev} \text{ survives} \\ p_{repair_m} \text{ if } acc_m \text{ survives and } C_{m\,prev} \text{ fails} \\ 0 \text{ if } acc_m \text{ fails and } C_{m\,prev} \text{ fails} \end{cases}$$

$$\mathrm{P}(C_m \text{ fails}) = \begin{cases} p_{f_m} \text{ if } acc_m \text{ survives and } C_{m\,prev} \text{ survives} \\ p_{f_m} \text{ if } acc_m \text{ fails and } C_{m\,prev} \text{ survives} \\ 1 - p_{repair_m} \text{ if } acc_m \text{ survives and } C_{m\,prev} \text{ fails} \\ 1 \text{ if } acc_m \text{ fails and } C_{m\,prev} \text{ fails} \end{cases} \quad (11)$$

3.8.3 Supply Components

In defining CPTs for supply components, the parent nodes of supply components typically include hazard nodes and service provision interdependency nodes. Equation (12) represents the CPT formulation for supply nodes. The supply node is denoted S_q, q = 1, …, n_S, where n_S is the number of supply nodes. The conditional probabilities of component failure given that a hazard occurs and does not occur are represented as $p_{f_{q|haz}}$ and

$p_{f_{q|no\ haz}}$, respectively. The hazard node of which the supply component is a child is represented by H_i and the service provision interdependency parent is represented by R_s.

$$\mathrm{P}\left(S_q \text{ survives}\right) = \begin{cases} 1 - p_{f_{q|haz}} \text{ if } H_i \text{ occurs and } R_s \text{ survives} \\ 0 \text{ if } H_i \text{ occurs and } R_s \text{ fails} \\ 1 - p_{f_{q|no\ haz}} \text{ if } H_i \text{ does not occur and } R_s \text{ survives} \\ 0 \text{ if } H_i \text{ does not occur and } R_s \text{ fails} \end{cases}$$

$$\mathrm{P}\left(S_q \text{ fails}\right) = \begin{cases} p_{f_{q|haz}} \text{ if } H_i \text{ occurs and } R_s \text{ survives} \\ 1 \text{ if } H_i \text{ occurs and } R_s \text{ fails} \\ p_{f_{q|no\ haz}} \text{ if } H_i \text{ does not occur and } R_s \text{ survives} \\ 1 \text{ if } H_i \text{ does not occur and } R_s \text{ fails} \end{cases} \quad (12)$$

3.8.4 Transshipment and Distribution Components

Transshipment and distribution components have MLSs and hazard nodes as parents. Equation (13) shows the CPT formulation for these transshipment and distribution components. The components are represented by the variable C_t, $t = 1, \ldots, n_d$, where n_d is the number of non-supply components. The MLSs that are parents of the component are represented as MLS_{d_v}, $v = 1, \ldots, n_M$, where n_M is the number of MLS parent nodes of component C_t. In equation (13), the component is also a child of a single hazard node denoted H_i. The conditional probabilities of component failure given that the hazard occurs and does not occur are represented as $p_{f_{t|haz}}$ and $p_{f_{t|no\ haz}}$, respectively.

$$P(C_t \text{ survives}) = \begin{cases} 1 - p_{f_{t|haz}} \text{ if } H_i \text{ occurs and any } MLS_{d_v} \text{ survives} \\ 1 - p_{f_{t|no\ haz}} \text{ if } H_i \text{ does not occur and any } MLS_{d_v} \text{ survives} \\ 0 \text{ if } H_i \text{ occurs and no } MLS_{d_v} \text{ survive} \\ 0 \text{ if } H_i \text{ does not occur and no } MLS_{d_v} \text{ survive} \end{cases}$$

$$P(C_t \text{ fails}) = \begin{cases} p_{f_{t|haz}} \text{ if } H_i \text{ occurs and any } MLS_{d_v} \text{ survives} \\ p_{f_{t|no\ haz}} \text{ if } H_i \text{ does not occur and any } MLS_{d_v} \text{ survives} \\ 1 \text{ if } H_i \text{ occurs and no } MLS_{d_v} \text{ survive} \\ 1 \text{ if } H_i \text{ does not occur and no } MLS_{d_v} \text{ survive} \end{cases} \quad (13)$$

3.8.5 Minimum Link Set Nodes

The functioning of an MLS depends on the functioning of the components in the MLS. Therefore, the parents of the MLS nodes are the components that comprise the MLS. However, there are two types of MLSs – those without cyclic links and those with cyclic links that have been removed. The formulations for the CPTs for the MLSs in the two cases are described below.

3.8.5.1 Non-Cyclic MLS Nodes

Equation (14) shows the CPT formulation for MLSs that did not contain cycles, and therefore do not contain links that have been removed with the cycle removal algorithm. The MLS nodes are denoted $MLS_w, w = 1, \ldots, n_{M_{NC}}$ where $n_{M_{NC}}$ is the number of non-cyclic MLSs. The components comprising the MLS are denoted $\mathbf{C}_{\mathbf{w}_x}, x = 1, \ldots, n_{w_x}$ where n_{w_x} represents the number of components in MLS_w.

$$\mathrm{P}(MLS_w \text{ survives}) = \begin{cases} 1 \text{ if all } \mathbf{C}_{\mathbf{w}_x} \text{ survive} \\ 0 \text{ if any } \mathbf{C}_{\mathbf{w}_x} \text{ fails} \end{cases}$$

$$\mathrm{P}(MLS_w \text{ fails}) = \begin{cases} 0 \text{ if all } \mathbf{C}_{\mathbf{w}_x} \text{ survive} \\ 1 \text{ if any } \mathbf{C}_{\mathbf{w}_x} \text{ fails} \end{cases} \tag{14}$$

3.8.5.2 Cyclic MLS Nodes

For MLSs containing cycles and therefore links that have been removed during the cycle identification process, the defined CPTs for these MLS nodes must account for the removed links. This is done using values from the joint probability distribution of the nodes for which links have been removed. Let the cyclic MLSs be denoted MLS_a, a = 1, …, n_{M_C} where n_{M_C} is the number of cyclic MLSs. The components that comprise an MLS are denoted $\mathbf{C}_{\boldsymbol{a_b}}$, b = 1, …, n_{a_b} where n_{a_b} represents the number of components in MLS_a. The removed links for each MLS are represented in $\mathbf{L_{rem}}$, a y ×2 matrix where y is the number of removed links for a specific MLS and each row represents the parent to child link that was removed. Each removed link is defined as $\mathbf{L}_{\mathbf{rem}_z}$, z = 1, …, y. The conditional probabilities of failure of components $\mathbf{C}_{\boldsymbol{a_b}}$ given that a hazard occurs and does not occur are represented as $p_{f_{b|haz}}$ and $p_{f_{b|no\ haz}}$, respectively. The probability of a hazard occurring is p_{H_i}, and the joint probability value calculated for use in the CPT is $p_{MLS\ cyc}$. The probability of failure of the link that is removed is calculated as the product of marginal failure probabilities of the parents of $\mathbf{L}_{\mathbf{rem}_z}$. The algorithm for formulating the CPT for MLSs with cyclic links is presented as **Algorithm C** follows:

Algorithm C. Algorithm for cyclic MLS CPT formulation.

for each row z in $\mathbf{L_{rem}}$:
Parents of $\mathbf{L_{rem_z}}$ are components corresponding to 1 values in adjacency matrix in row $\mathbf{L_{rem_z}}$

$$\mathrm{P}(\mathbf{L_{rem_z}}\text{ fails}) = \prod_{b=1}^{n_{a_b}} \left[p_{f_{b|haz}} \cdot p_{H_i} + p_{f_{b|no\ haz}} \cdot (1 - p_{H_i})\right]$$

$$\begin{aligned} p_{MLS\ cyc} = {} & \sum_{z=1}^{y} \mathrm{P}(\mathbf{L_{rem_z}}\text{ fails}) - \sum_{1\le c<d\le y} \mathrm{P}(\mathbf{L_{rem_c}}\text{ fails}) \cap \mathrm{P}(\mathbf{L_{rem_d}}\text{ fails}) \\ & + \sum_{1\le c<d<f\le y} \mathrm{P}(\mathbf{L_{rem_c}}\text{ fails}) \cap \mathrm{P}(\mathbf{L_{rem_d}}\text{ fails}) \cap \mathrm{P}(\mathbf{L_{rem_f}}\text{ fails}) - \cdots \\ & + (-1)^{n-1}[\mathrm{P}(\mathbf{L_{rem_1}}\text{ fails}) \cap \ldots \cap \mathrm{P}(\mathbf{L_{rem_y}}\text{ fails})] \end{aligned}$$

$$\mathrm{P}(MLS_a\text{ survives}) = \begin{cases} 1 - p_{MLS\ cyc} \text{ if all } \mathbf{C_{a_b}} \text{ survive} \\ 0 \text{ if any } \mathbf{C_{a_b}} \text{ fails} \end{cases}$$

$$\mathrm{P}(MLS_a\text{ survives}) = \begin{cases} p_{MLS\ cyc} \text{ if all } \mathbf{C_{a_b}} \text{ survive} \\ 1 \text{ if any } \mathbf{C_{a_b}} \text{ fails} \end{cases}$$

The definition of joint probability, $p_{MLS\ cyc}$, scales exponentially. However, with the example 112-component network described in 4.2, the calculation of CPTs for the 392 nodes in the BN takes only one minute. Once the CPTs for all nodes are defined, the BN model can be built.

3.9 Proposed Combination Metrics

In analyzing methods for assessing component importance, centrality-based metrics are useful in that they account for connectivity of a network when calculating component criticality. However, they do not consider individual component reliability, which affects quantification of the impacts of individual components on overall system performance. Therefore, centrality-based metrics should not be used alone to rank component importance. Reliability-based metrics do include component reliability. However, they give less weight to the topology of the network compared to the centrality-based metrics discussed. In assessing component criticality in interdependent infrastructure networks, both system-level topology and component-level reliability characteristics are important. Therefore, two measures that combine centrality-based and reliability-based attributes are proposed.

Previously, Cadini, Zio, and Petrescu (2009) proposed four reliability centrality measures for network infrastructure and applied these to a power transmission system. The measures are reliability degree centrality, reliability closeness centrality, reliability betweenness centrality, and reliability closeness centrality. These combine the classical definition of the centrality metric with a reliability metric such as edge reliability or path reliability. The reliability degree centrality, for example, is calculated by multiplying a node's degree by the reliability of its connected links (Cadini, Zio, & Petrescu, 2008, October).

Compared to the previous study, this dissertation proposes accounting for component reliabilities in the approach. This is due to the importance of individual component performance in governing system performance, i.e., the probability of providing an infrastructure service at a final distribution point will be dependent on the reliabilities of individual component comprising the system. In this study, the importance of node i is calculated including its reliability with weighting by centrality. The combination metric proposed, W_i^{AD}, is calculated using equation (15):

$$W_i^{AD} = W_i^A \cdot w_i^D \tag{15}$$

where W_i^A is RAW and w_i^D represents the weight for the component based on its degree. The weight used for the combination metric is the normalized degree value plus one and is calculated using equation (16). One is added so that the components with a single connection will have a weight of one rather than a weight of zero.

$$w_i^D = 1 + \frac{C_i^D - \min(C^D)}{\max(C^D) - \min(C^D)} \tag{16}$$

A second combination metric is offered (W_i^{AM}) using MLS appearances and RAW. This is calculated using equation (17):

$$W_i^{AM} = W_i^A \cdot w_i^M \tag{17}$$

where w_i^M represents the weight for the component based on its number of appearances in MLSs defined in the system using **Algorithm A**. This weight is the normalized number of MLS appearances plus one and is calculated using equation (18).

$$w_i^M = 1 + \frac{C_i^M - \min(C^M)}{\max(C^M) - \min(C^M)} \tag{18}$$

These combined measures enable the accounting for both topological and individual component reliability characteristics when assessing component criticality.

CHAPTER 4. APPLICATION EXAMPLE FOR PROPOSED FRAMEWORK

4.1 Introduction

To demonstrate the proposed framework and approach, it is applied to the interdependent water and power distribution networks in Atlanta, Georgia, in this chapter. The system is modeled, and inference is performed on the network using the model. The chapter includes a description of how each step described in Chapter 3 is applied to the example network. The chapter starts with an overview of the system studied, the inputs supplied, the dimensionality reduction, and the construction of the BN. The computational requirements in terms of memory storage and computation time to run several of these steps are included.

4.2 System Overview

For the water system, pipes greater than or equal to 18 inches in diameter are analyzed. This includes 112 components, seven of which are supply stations and 105 of which are transshipment or distribution nodes. There are 244 links, or pipes, in the network. For the power system, the power substations that are located at each supply node are modeled. Supply nodes have between one and three electrical feeds, varying with each supply component.

Figure 6 shows the system with supply nodes shown as empty circles and distribution and transshipment nodes shown as solid points. The supply nodes are also the locations of the power components.

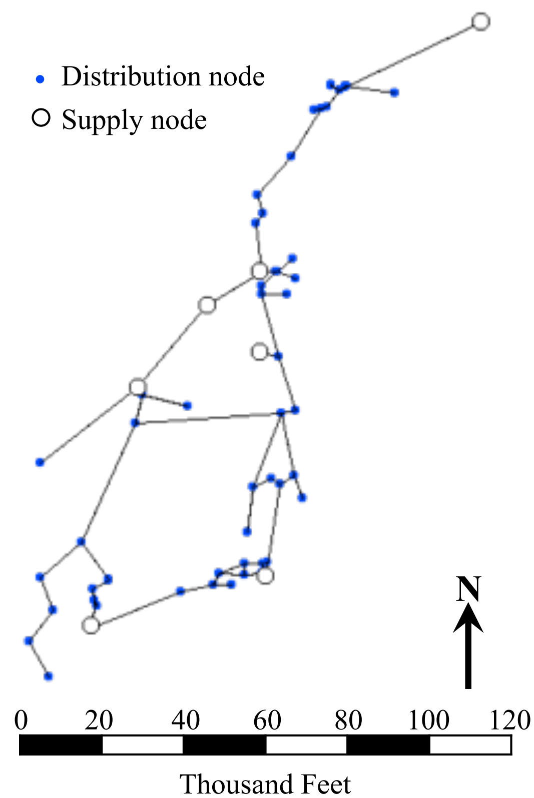


Figure 6. Atlanta, Georgia water and power distribution systems.

4.3 Inputs

Inputs include identification numbers and locations of 112 junctions in the water network. The start and end junctions and size characteristics of 350 pipes are included. The junctions are condensed to represent the start and end junctions of each pipe rather than accounting for all on-pipe junctions.

4.3.1 Component Locations

The component locations are given as state-plane coordinates. An example of locations for two components C1 and C2 is shown in Table 3.

Table 3. Component locations for example system in state-plane coordinates.

Node	Latitude	Longitude
C1	2.2452e6	1.4471e6
C2	2.1893e6	1.3204e6

4.3.2 Component Connectivity

The component connectivity for the application is obtained from a list of each link in the network used in the hydraulic model of the system. Table 4 is an excerpt from this list of connections, where the start and end nodes are names of junction components in the network.

Table 4. Component connectivity for example system.

Start	End
C1	C79
C2	C33
C2	C38

4.3.3 Component Type

The component types for the application are defined depending on their function, i.e., supply, transshipment, or distribution. The constituent elements of supply nodes, i.e., for pump stations and treatment plants, are aggregated into a single node for each supply.

4.3.4 *Component Probability of Failure*

For the application, component failure probabilities are assumed to be consistent across each component to better assess relative component vulnerabilities. Two levels of operation are considered – with occurrence of a hazard and without occurrence of a hazard. The failure probabilities given that the hazard occurs or does not occur are assumed to be 1×10^{-2} or 1×10^{-4}, respectively. The hazard in the example is generalized and could, e.g., represent a storm. The equal prior failure probabilities across components results in ranking and component prioritization rather than obtaining specific failure probability values. If more information is learned about the components, the failure probabilities can be easily updated as inputs to the model.

4.4 Dimensionality Reduction

Running **Algorithm A** for the full system identifies the MLSs from a supply node to each of the transshipment and distribution nodes in the network. This takes approximately 2.19 seconds on a computer with 4 GB RAM and 1.3 GHz Intel Core i5 processor using MATLAB 2017b for the entire network. There are 246 MLSs in the full system. The maximum number of MLSs for a component is five components and the maximum length of an MLS is 17 components. An example set of MLSs for node C7 is:

$$\begin{bmatrix} \text{C108, C58, C59, C7} \\ \text{C108, C60, C59, C7} \end{bmatrix}$$

where the first component in each row is a supply node and the middle nodes are on the path to the final node. Super-components were not needed for this example.

4.5 Defining Interdependencies

The interdependencies modeled in the application are service provision and geographic. Service provision interdependencies are based on information provided by the owners of the water network. There are power substations located at each of the water supply stations. To model the service provision interdependencies, direct links are added from each power substation to the water supply node that it supplies. Backup generators can also be incorporated to account for continued power in the case of an outage of a main substation. There is a total of seven power substations in the network at the same locations as their corresponding water supply components that provide power to seven water supply nodes.

The water and power networks are partitioned into hazard zones that are used to represent geographic interdependencies. These hazard zones also represent service areas surrounding each of the water supply nodes. The seven zone partitions for the network are shown in Figure 7.

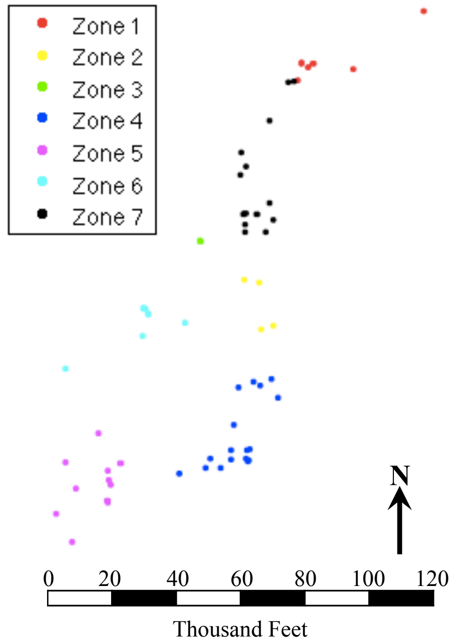


Figure 7. Service areas partitioned by zone for application network.

Two of the service areas are split into two groups heuristically for ease of computation during inference. Therefore, in total, there are nine partitions with hazard nodes as parents for the nodes in each of them.

4.6 Interdependent Adjacency Matrix Construction

The next step is to build the interdependent adjacency matrix from identified MLSs and interdependency relationships between nodes. Each MLS is a parent of its dependent component node, and the components that comprise the MLS are parents of the MLS node. Links created by service provision interdependencies include defined parent and dependent children nodes. Geographic interdependency links include those from a hazard node to each component in the hazard zone, as well as links from each distribution component in a zone to the service level nodes used for posterior inference. There are 1,868 parent-child relationships defined in the adjacency matrix for the network of 112 components and 244 links. The Bayes Nets Toolbox (Murphy, 2001) is used to construct and perform inference in MATLAB. This toolbox requires the components to be sorted topologically from parent to child nodes. The sorting and construction of the adjacency matrix takes approximately 1.26 seconds.

4.7 Accounting for Cycles

Defining all MLSs and parent-child relationships for the example system creates 52 cycles in the network. A graph traverse algorithm is applied to first identify these cycles and then remove one of the links from a component to a MLS for each cycle. This removes the cycles from the network in a negligible amount of time.

4.8 Defining Conditional Probability Tables

Overall, the definition of conditional probabilities takes about 62.6 seconds. This builds CPTs for each of the 392 nodes in the BN. The minimum CPT size is 2×1 for hazard nodes. For nodes with n parents, the size of the CPT is $2 \times 2 \times (n + 1)$.

4.8.1 Geographic Interdependency Nodes

The probability that a hazard occurs is assumed to be 0.01. This is for a generalized hazard and can be changed to the probability of occurrence of any specific hazard of concern. As an example, the CPT for service zone 1 is:

$$P(Z_1 \text{ is at N\% Service}) = \begin{cases} 1 \text{ if N\% of } C_{Z_j}^k, k = 1, \dots, 12 \text{ survive} \\ 0 \text{ otherwise} \end{cases}$$

where $C_{Z_1} = C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19$

C_{Z_1} includes all transshipment and distribution components in zone 1. The CPT in the above example is $2 \times 2 \times 13$.

4.8.2 Supply Components

The CPT for an example supply component, C108, is shown in Table 5. In Table 5, H_9 indicates the hazard in its zone partition and R_1 and R_2 are the two power substations that supply C108. S and F denote survival and failure, respectively, of the substations.

Table 5. CPT for example supply component, C108.

C108	H_9 Occurs				H_9 Does not occur			
	R_1 S		R_1 F		R_1 S		R_1 F	
	R_2 S	R_2 F	R_2 S	R_2 F	R_2 S	R_2 F	R_2 S	R_2 F
Survives	0.99	0.99	0.99	0	0.9999	0.9999	0.9999	0
Fails	0.01	0.01	0.01	1	0.0001	0.0001	0.0001	1

4.8.3 Transshipment and Distribution Components

The CPT for an example transshipment or distribution component, C99, is shown in Table 6. In Table 6, H_3 indicates the hazard in its zone partition. For this example, component C99 has only one MLS parent. S and F denote survival and failure, respectively, of the MLS.

Table 6. CPT for example distribution component, C99.

C99	$\mathrm{H_3}$ Occurs		$\mathrm{H_3}$ Does not occur	
	$\mathrm{MLS_{C99_1}}$ S	$\mathrm{MLS_{C99_1}}$ F	$\mathrm{MLS_{C99_1}}$ S	$\mathrm{MLS_{C99_1}}$ F
Survives	0.99	0	0.9999	0
Fails	0.01	1	0.0001	1

4.8.4 Minimum Link Set Nodes

An example of the CPT for a non-cyclic MLS, $\mathrm{MLS_{C70_1}}$, is shown in Table 7, where $\mathrm{C_{MLS_{C70_{1_1}}}}$ is C111 and $\mathrm{C_{MLS_{C70_{1_2}}}}$ is C91; these are the two nodes that comprise $\mathrm{MLS_{C70_1}}$.

Table 7. CPT for example non-cyclic MLS, $\mathbf{MLS_{C70_1}}$.

$\mathbf{MLS_{C70_1}}$	$\mathrm{C_{MLS_{C70_{1_1}}}}$ S		$\mathrm{C_{MLS_{C70_{1_1}}}}$ F	
	$\mathrm{C_{MLS_{C70_{1_2}}}}$ S	$\mathrm{C_{MLS_{C70_{1_2}}}}$ F	$\mathrm{C_{MLS_{C70_{1_2}}}}$ S	$\mathrm{C_{MLS_{C70_{1_2}}}}$ F
Survives	1	0	0	0
Fails	0	1	1	1

An example of the CPT for a cyclic MLS, $\mathrm{MLS_{C66_3}}$, is shown in Table 8. In this example, the link that is removed is from C3 to $\mathrm{MLS_{C66_3}}$. Therefore, the remaining parent of $\mathrm{MLS_{C66_3}}$ is $\mathrm{C_{MLS_{C66_{3_1}}}}$, which represents C110. The probability that C3 survives is calculated using **Algorithm C**.

Table 8. CPT for example cyclic MLS, $\mathbf{MLS_{C66_3}}$.

$\mathbf{MLS_{C66_3}}$	$C_{MLS_{C66_{3_1}}}$S	$C_{MLS_{C66_{3_1}}}$F
Survives	P(C3 Survives)	0
Fails	1 – P(C3 Survives)	1

4.9 Output

Figure 8 shows the overall BN model. The hazard nodes are denoted Hazard_1, …, Hazard_9. These are parents of the power and water components and represent the geographic interdependency. The power supply components are denoted Power Supply_1, …, Power Supply_7. These are parents of water supply components – representing service provision connections. Water supply components are denoted Water Supply_1, …, Water Supply_7, and water distribution components $\text{Water Distribution}_1$, …, $\text{Water Distribution}_{105}$. Water distribution components are parents of zones partitions Zone_1, …, Zone_9, which represent levels of service throughout the network. Both water supply and distribution components are parents of MLSs denoted Water MLS_1, …, Water MLS_{246}. MLSs are parents of the distribution components that they supply. The subscripts represent the number of nodes of each type in the network; the BN comprises 383 total nodes.

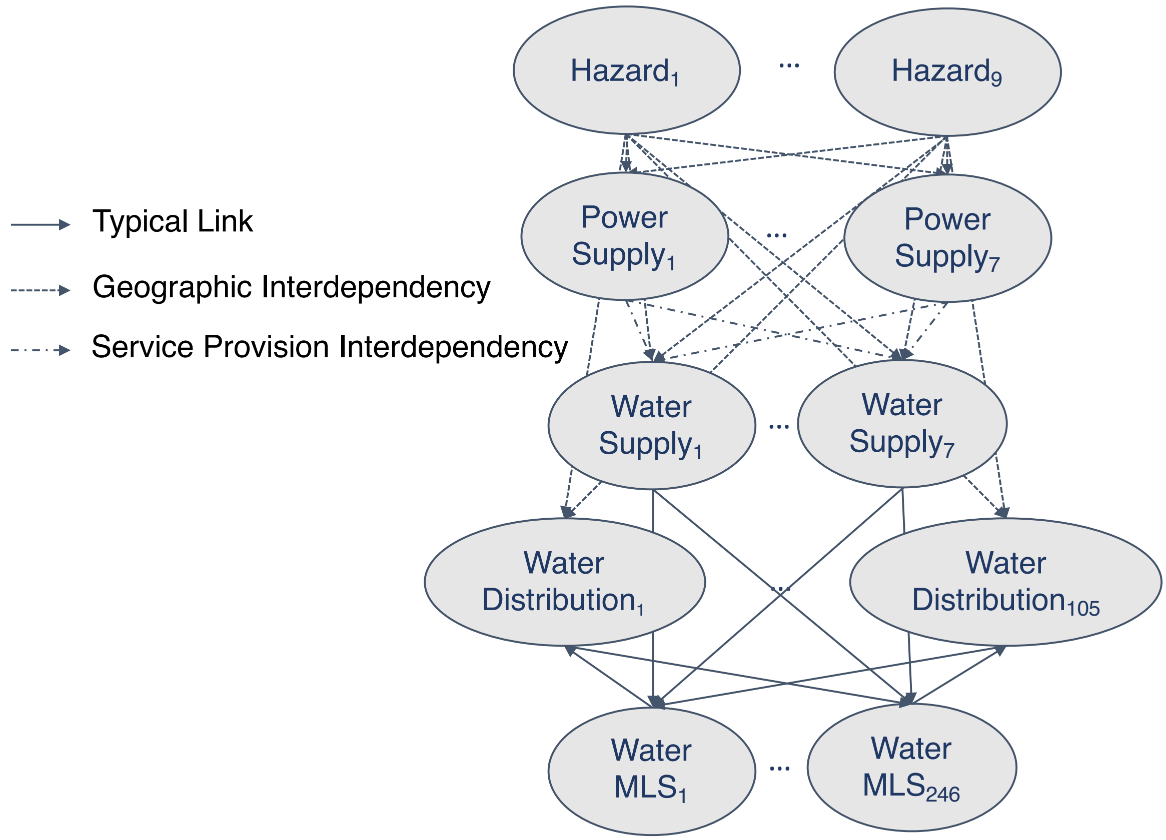


Figure 8. Overall BN model of Atlanta water and power distribution networks.

CHAPTER 5. VALIDATION FOR INTERDEPENDENT INFRASTRUCTURE NETWORKS WITH APPLICATION

5.1 Introduction

This chapter provides a validation of the use of the framework including both external and internal validation using the model of Atlanta's infrastructure networks as the illustrative example. The chapter starts with external validation including assessing the accuracy of the model using a real-world scenario. Then, internal validation is provided with other applications of the interdependent BN model and example inferences for specific scenarios.

5.2 External Validation

The accuracy of the model is validated using a real-world scenario of cascading failures due to the interdependent nature of infrastructure networks that occurred in both 2014 and 2017. In these instances, a water pump station lost power from both of its dual feeds and caused outages throughout Atlanta's downtown area. The water system lost pressure in both cases and a boil water advisory became necessary. To test the scenario with the model, an outage was simulated to the power components supplying the affected pump station. The resulting network showed outages throughout the downtown area, as shown in Figure 9. This is consistent with the outcomes of the event where the downtown area lost water pressure. The loss of water pressure is used as an indicator for failure at the distribution level in the example. The BN model includes the complexities of the functionality and interdependencies of the networks and shows the effects of the outage directly.

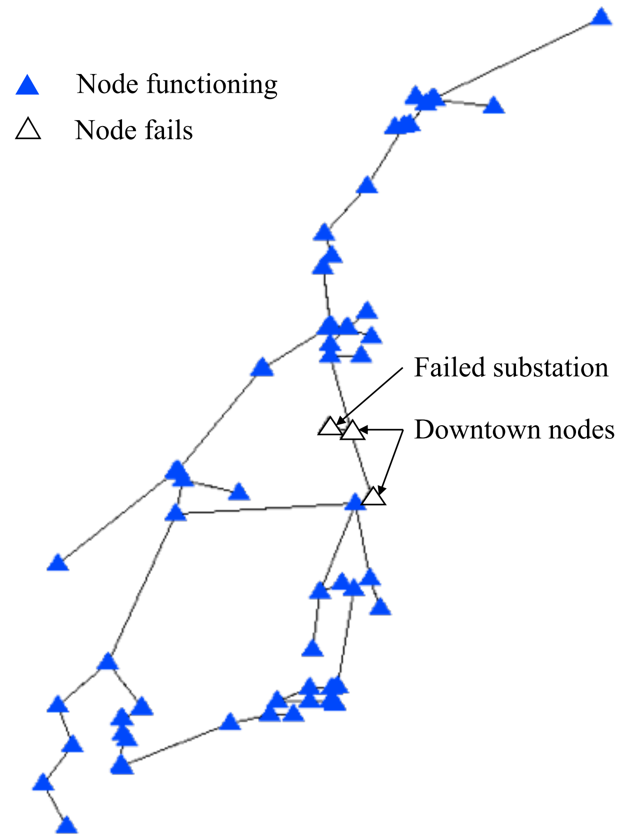


Figure 9. Atlanta outage scenario for validation.

5.3 Internal Validation

With the BN model built, varying inferences can be conducted over the networks. The external validation scenario above is an example of assessing the impacts of a service provision interdependency, where the power supply of a water pump station failed and caused cascading outages in the water system. Other inference examples to perform probabilistic vulnerability analyses include assessing the impacts of a hazard occurring in a specific zone – a geographic interdependency – or evaluating the effects of failures within the water system itself.

Figure 10 shows inference results from a hazard occurring in hazard zones 1 and 2. The gradient on the right represents failure probabilities. Hazard zones 1 and 2 are in the upper right corner of the system, so it is observed that components in that area experience increased probabilities of failure. As the supply nodes are distributed throughout the rest of the network, no additional outages are experienced due to this event scenario.

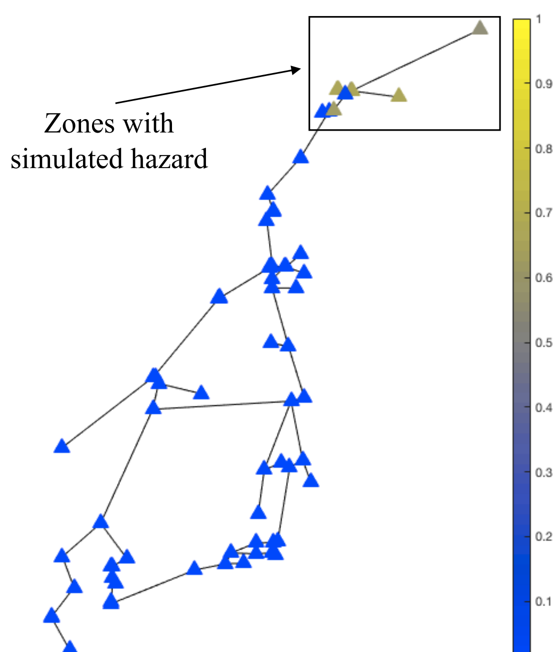


Figure 10. Inference results from hazard occurrence in zones 1 and 2.

Another example of inference is to assess the effects of an observed outage or failure of a specific component in the network. Inference over the BN will update the failure probabilities of all nodes throughout the network. Figure 11 shows the results from learning that a large supply component in the bottom right area of the network has failed. The effects of such an outage depict the ability to provide service in that part of the network.

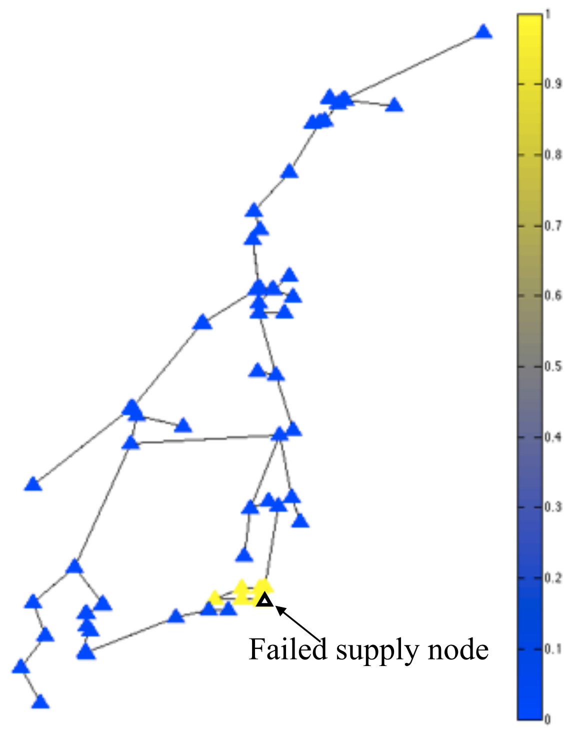


Figure 11. Inference results from supply node failure.

These inferences performed highlight the abilities of the proposed framework. The results shown are a small subset of information that can be gained from the built interdependent infrastructure model. The models allow a user to input information across a wide range of possible scenarios, e.g., outages that are experienced or expected; hazard occurrences; or updated information on a component such as failure, retrofit, or replacement. The user can then visualize and observe the updated probabilities of failure in components throughout the network. The above inferences were performed in approximately 4 seconds each. The output is achieved in a computationally efficient manner and is based on a full representation of the network, including the performance of its constituent individual components and the interdependencies that exist across systems.

CHAPTER 6. COMPONENT IMPORTANCE MEASURES WITH APPLICATION

6.1 Introduction

This chapter provides an overview of the centrality-based, reliability-based, and combination component importance measures that are applied to the example Atlanta water distribution system and its dependencies on power. A comparison of the five metrics is provided in Section 6.3, as well as an assessment of the impact that considering dependencies when ranking component criticality has in Section 6.4.

6.2 Critical Component Identification

To demonstrate the proposed component importance measures, they are applied to the interdependent water and power distribution networks in Atlanta, Georgia. Degree centrality, MLS appearances, RAW, and the two combination metrics are applied to the water system alone and to the interdependent water and power systems. Section 6.4 includes a discussion of the effect of considering interdependencies on the component importance rankings.

6.2.1 Centrality-Based Metric – Degree Centrality

Degree centrality for the networks is calculated by summing the number of connections for each component. More connections represent nodes with higher degree thus, lower rankings and higher criticality. The water system is first analyzed on its own. In this analysis, the top three most critical nodes are water supply components. Specifically, supply nodes S3, S5, and S6 are each connected to five other nodes in the network. The system is then analyzed including the dependencies on power to see how the rankings change. The water supply components increase in importance with a higher maximum degree now of six, with supply components S3, S5, and S6 remaining the most critical. In comparison, fifty nodes have a degree of one, including all of the power nodes.

The system is then analyzed without the power components to see how rankings change without consideration of interdependencies. In the analysis, water supply components decrease in importance with a lower maximum degree now of five. Supply components S3, S5, and S6 have a degree of five and remain the most critical.

To assess the impact of interdependencies on rankings of component criticality, Figure 12 and Figure 13 show the degree rankings with and without consideration of interdependencies. Figure 12 plots the rankings with consideration of interdependencies as squares and without consideration of interdependencies as circles on the y-axis. The components are ordered on the x-axis by their ranking when interdependencies are considered. The background of the plot for each component indicates the component type. The white background represents water distribution nodes. The lightest gray represents water transshipment nodes. The darker gray represents water supply nodes. The darkest background represents power components. Components with lower rankings (i.e., 1-5) are more critical than components with higher rankings. Figure 13 displays the component rankings on the layout of the network. The darkest color represents the most critical nodes. Figure 13(a) shows rankings with consideration of interdependencies and Figure 13(b) represents rankings without consideration of interdependencies. In some cases, more critical nodes are covered by less critical nodes, such as S3 and S6 in Figure 13(a). While the components appear a lighter color, they are covered by a less critical node in the plot.

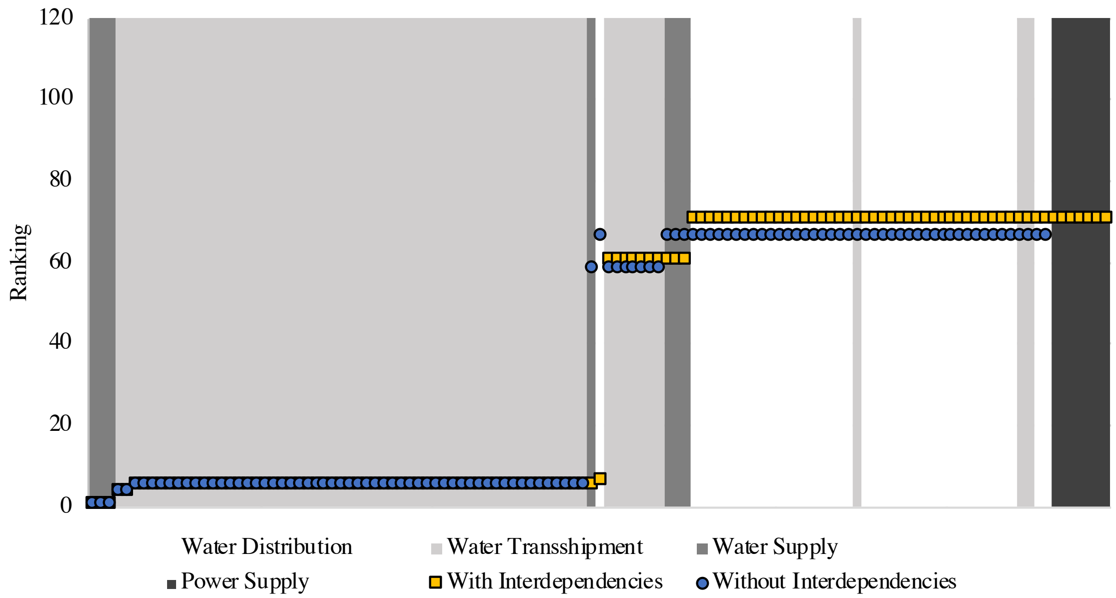


Figure 12. Degree rankings with and without consideration of interdependencies.

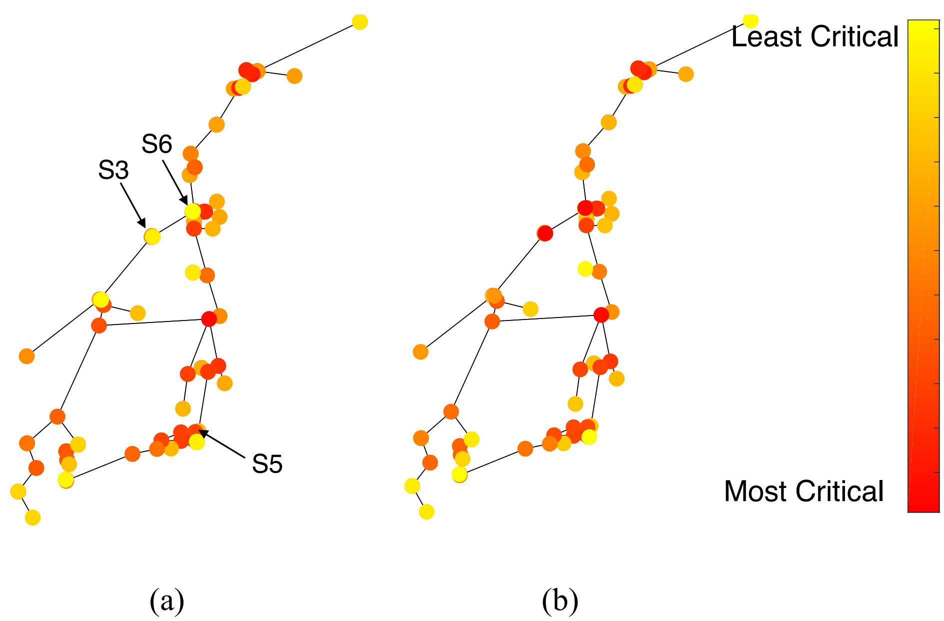


Figure 13. Network showing degree rankings with (a) and without (b) consideration of interdependencies.

In calculating degree centrality, many components have a degree of three or one in the network. This is typical for infrastructure networks designed in a branching topological structure or for end-point distribution nodes, respectively. Therefore, as seen in Figure 12, many components have a similar ranking. In fact, for the entire network, there are only five different ranking levels. Thus, degree centrality should not be used on its own to prioritize infrastructure investment, as it is necessary to further distinguish between component criticality beyond a few coarse levels.

6.2.2 Centrality-Based Metric – MLS Appearances

The number of MLS appearances for a component is calculated by summing the number of MLSs for which each node is a part. Lower rankings represent nodes with more MLS appearances, and are therefore more critical. The maximum number of MLS appearances is 74 and the minimum is zero. As MLSs represent flow from supply nodes to distribution nodes, power components are not included in the MLSs, so have the lowest importance based on MLS appearances. For this reason, the rankings without consideration of interdependencies are the same as those with consideration of interdependencies, excluding the power components. Figure 14 and Figure 15 show the rankings using the number of MLS appearances ordered by component importance and on the layout of the network, respectively.

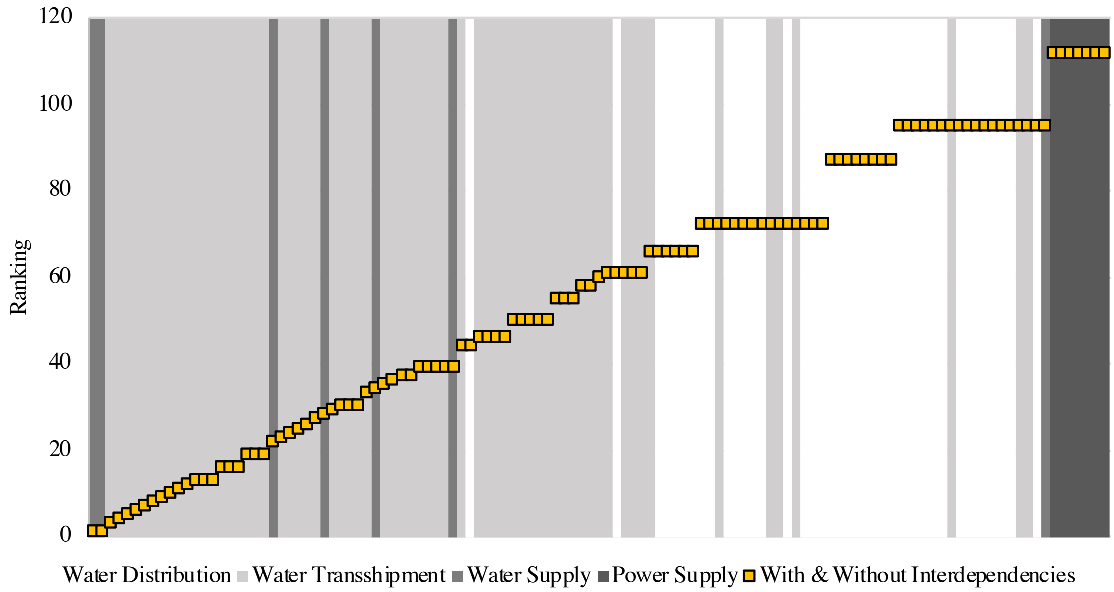
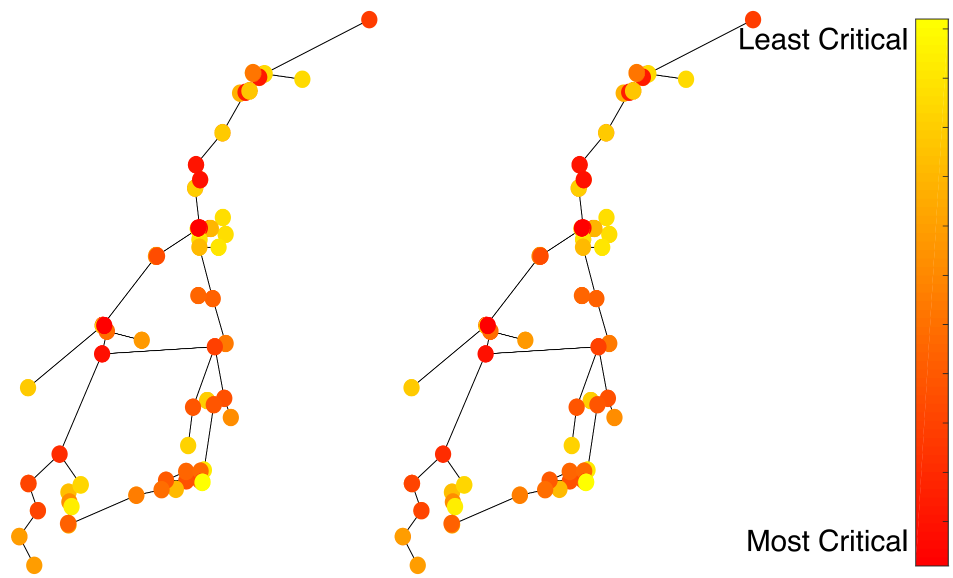


Figure 14. MLS Appearances rankings with and without consideration of interdependencies.



(a) (b)

Figure 15. Network with MLS Appearances rankings with (a) and without (b) consideration of interdependencies.

From Figure 14, number of MLS appearances provides a larger variation in component importance rankings than degree centrality. However, with only the consideration of number of flow paths from supply components to distribution components, these rankings do not consider dependencies on power for the water system or the reliability of individual components. Therefore, this measure cannot be used on its own to assess component criticality in interdependent infrastructure networks

6.2.3 Reliability-Based Metric – Risk Achievement Worth

The interval-scale RAW measure is used to rank component criticality including individual component reliabilities. First, prior probabilities of failure are calculated for each component based on the BN model constructed. Then, inference is performed to calculate each RAW value by updating the network with a 100% probability of failure for that component. The probabilities of failure of every other component are then calculated. An aggregate RAW is calculated by summing the changes in failure probability for each component. For example, the RAW of component D1 is calculated by indicating that component D1 has a 100% probability of failure. The updated probabilities of failure of all components (D2 to D105, S1 to S7, and P1 to P7) are then calculated using the BN. The changes in probabilities of failure are then calculated by subtracting the new probabilities of failure from the prior probabilities of failure for each component. The final RAW of component D1 is the sum of the changes in probabilities of failure of all other components. Components with lower rankings have a higher RAWs have lower rankings and therefore higher criticality.

Figure 16 plots the rankings from RAW with and without consideration of interdependencies. Components are ordered by their rankings with interdependencies. Figure 17(a) shows the rankings from RAW with consideration of interdependencies on a graph of the network. Again, the darker circles represent higher criticality. Dispersion represents the change in ranking for a component with and without consideration of dependence on power. Figure 17(b) shows the rankings of components using RAW without consideration of interdependencies.

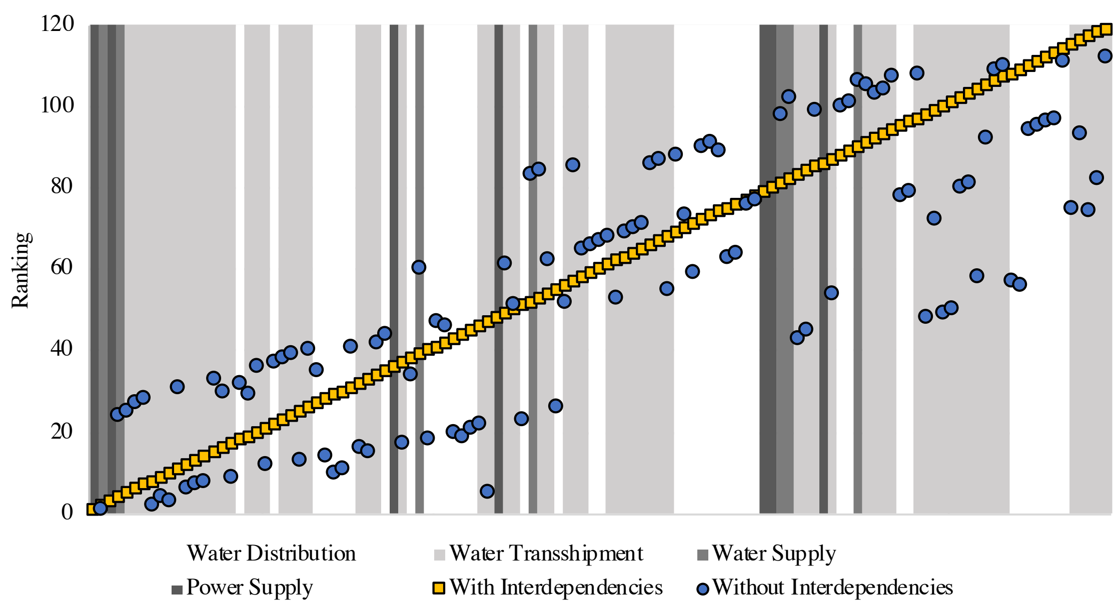
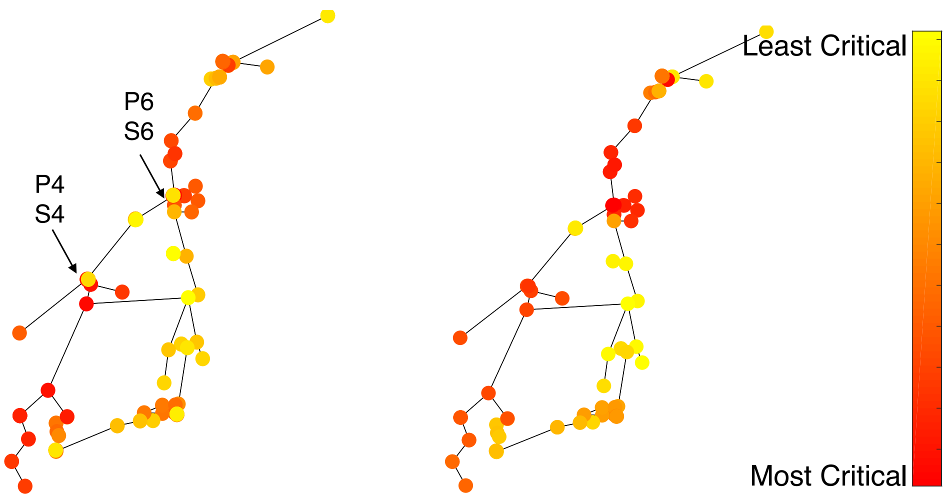


Figure 16. RAW rankings with and without consideration of interdependencies.



(a) (b)

Figure 17. Network with RAW rankings with (a) and without (b) consideration of interdependencies.

The top four most critical components when dependencies on power are considered are water supply components S4 and S6 and their corresponding power supplies, P4 and P6. When RAW rankings are calculated without consideration of dependencies on power, water supply S6 is the most critical node, but S4 has moved to the 24th most critical node. The RAW value of both S6 and S4 are similar with and without consideration of dependence on power. However, there is a small range in RAW values when dependence on power is not considered. The maximum difference in RAW values between the top 24 most critical components is approximately 9%. The other nodes with the highest criticality when dependence on power is not considered are transshipment nodes. The second most critical node becomes D38, with an increased RAW value of approximately 50% from the interdependency case. Each of the transshipment nodes that are the second to ninth most critical without consideration of dependence on power are relatively close to supply components: six of the components are one link away from a supply, one is two links away, and one is three links away. Since updates to a component state in the BN model propagate both to the children and parents of the node, evidence of failure of each of these transshipment components increases the probability of failure of the supply nodes included in the MLSs. Here, RAW captures the secondary effects from supply to transshipment nodes compared to the importance of the supply nodes themselves as in the interdependency case. In these rankings, RAW is useful in that it accounts for component-level reliabilities. However, RAW gives a less direct weight on the importance of topology in component importance rankings. It is therefore recommended that RAW be combined with a centrality-based metric to directly account for topology as well.

6.2.4 Proposed Combination Metrics

Considering the limitations of using degree centrality, MLS appearances, or RAW on their own to assess component importance, two combination metrics are proposed and analyzed to include both centrality-based and reliability-based attributes.

6.2.4.1 Degree and RAW Combination Metric

The first combination metric assessed is combining degree with RAW both with and without consideration of interdependencies. Figure 18 and Figure 19 show the rankings using the degree and RAW combination metric.

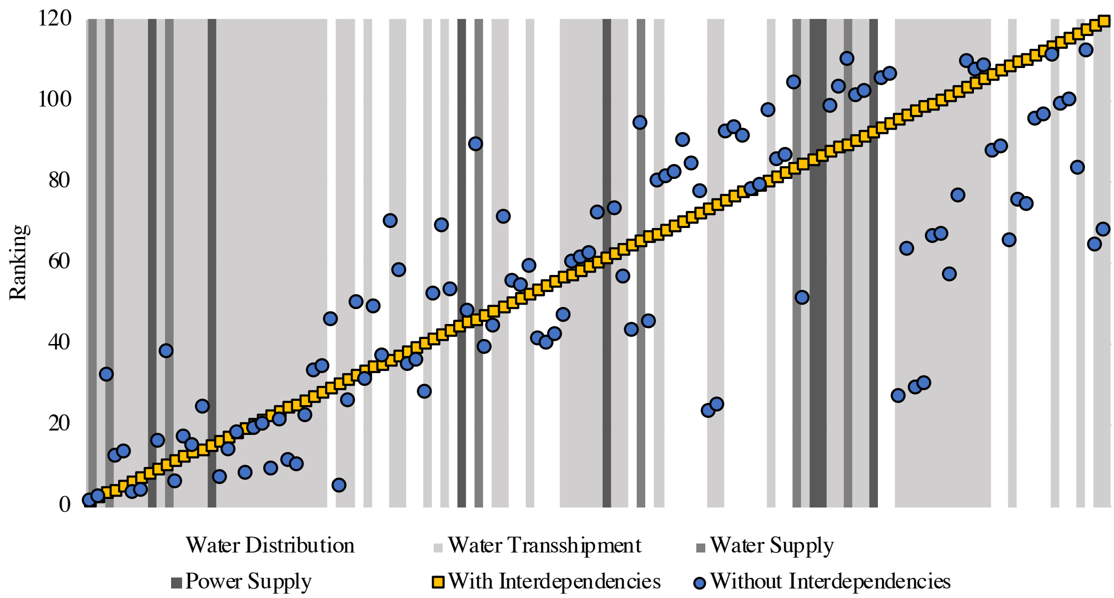


Figure 18. Degree + RAW rankings with and without consideration of interdependencies.

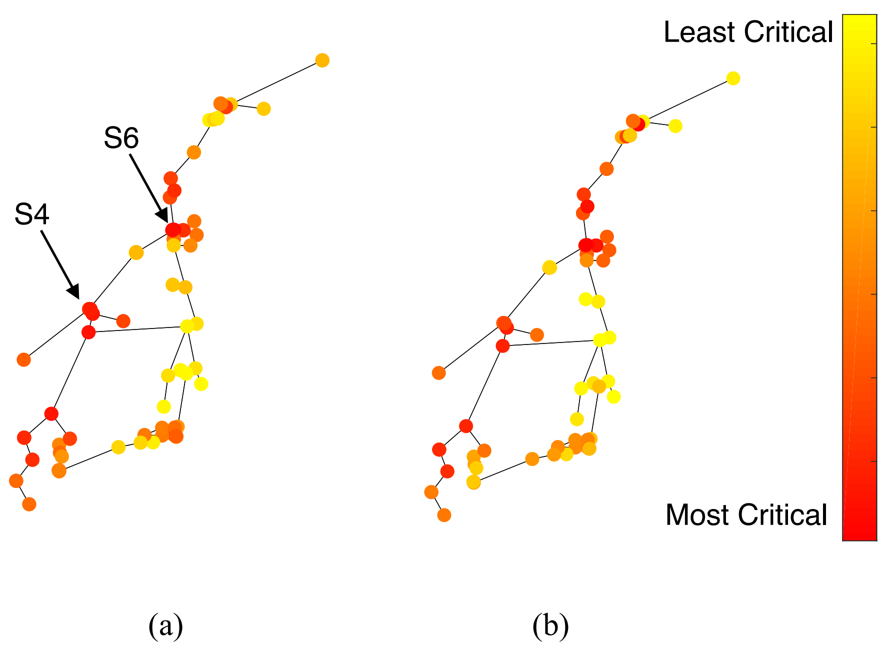


Figure 19. Network with Degree + RAW rankings with (a) and without (b) consideration of interdependencies.

Two of the most critical components when dependencies on power are considered are supply nodes S5 and S6. However, S4 moves from third to 54th most critical when dependence on power is not considered. S6 is the most critical component in rankings using both degree and RAW with and without consideration of dependence on power, showing consistent results with the individual degree and RAW rankings. Power supply components range from the eighth most critical (P6) to the 92nd most critical (P1). All of the water supply components are more critical when dependencies on power are considered compared to when they are not, with the exception of S6, which remains the most critical. This is also consistent with the RAW rankings individually. As supply nodes are necessary to deliver resources to transshipment and distribution nodes, this supports the recommendation in this study to consider interdependencies when prioritizing maintenance, repair, and retrofit decisions.

6.2.4.2 MLS Appearances and RAW Combination Metric

The last metric assessed is weighting RAW by the number of MLSs in which each component appears with and without consideration of dependencies on power. Figure 20 and Figure 21 show the ranking of components with and without consideration of interdependencies using the MLS appearances and RAW combination component importance metric.

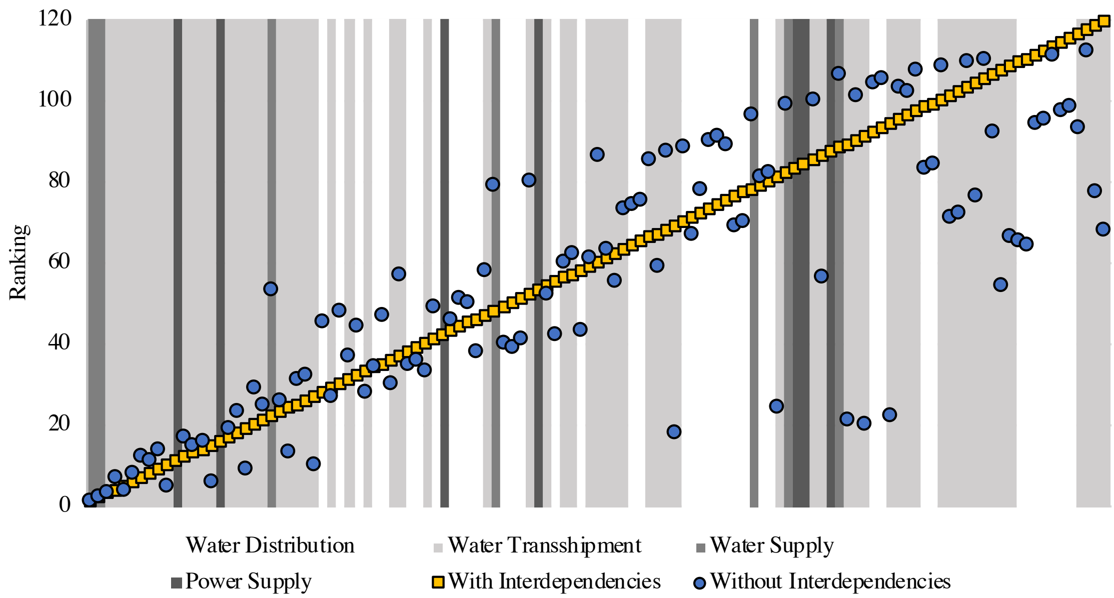


Figure 20. MLS Appearances + RAW rankings with and without consideration of interdependencies.

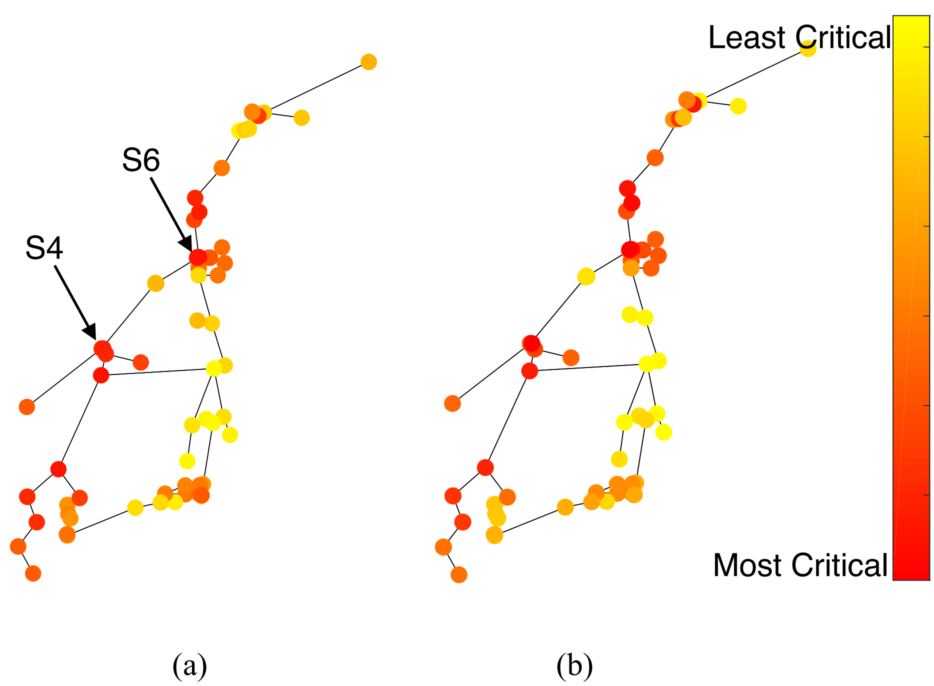


Figure 21. Network with MLS Appearances + RAW rankings with (a) and without (b) consideration of interdependencies.

Components S4 and S6 are the most critical based on this metric when dependencies on power are and are not considered. S4 does not move down in rankings as with RAW alone when dependence on power is removed because it is one of the most critical components by MLS appearances. Even though they appear in no MLSs, the relatively high RAW values of the power components mean that they maintain their criticality in the combination metric rankings. If more information were available about the power system and its connectivity, this combination metric would be the most appropriate because it accounts for component reliability and MLS appearances representing a wide range of component connectivity. I.e., information on the power network would account for redundancies in power supply to the water network, capturing criticality within the power system while not creating redundancies with the water supply components.

6.3 Comparison of Metrics

In the previous sections, the outcomes of component criticality rankings for the connected water and power networks in Atlanta were shown. Component importance was assessed using three individual metrics and two combination metrics. Intuitively, water supply nodes should be more critical than distribution nodes, e.g., distribution nodes will not be able to provide service if supplies are out. In the assessment of infrastructure interdependencies, the connections between the water and power nodes are critical. Therefore, to assess how the types of nodes vary in assessed importance based on the measure used, Table 9 shows a summary of the results based on component type of water supply nodes and power supply nodes. The top five most critical nodes using each metric are also provided. The transshipment and distribution nodes are ranked in the top five using each of the metrics are in close proximity to supply components, as described in section

6.2.3. The failure of these nodes increases the updated probabilities of failure of their surrounding supply components.

Table 9. Comparison of component importance metrics with consideration of interdependencies.

	Degree	MLS Appearances	RAW	Degree + RAW	MLS + RAW
Distribution of Water Supply	• Top 3 most critical • Top 58%	• Top 2 most critical • Top 38%	• Two in top 4 most critical • Top 76%	• Two in top 3 most critical • Top 75%	• Two in top 3 most critical • Top 75%
Distribution of Power Supply	• Seven least critical	• Seven least critical	• Two in top 3 most critical • Top 72%	• Between top 7% and 77%	• Between top 9% and 75%
Top 5 Most Critical	• S3 • S5 • S6 • D60 • D104	• S4 • S6 • D60 • D61 • D38	• P6 • S6 • P4 • S4 • D60	• S6 • D60 • S4 • D61 • D63	• S6 • S4 • D60 • D61 • D38

Water and power supply components are necessary to provide water throughout the remaining components in the network. Therefore, the results for these components are used as a means to make recommendations for a useful and accurate metric to rank component criticality. The two centrality-based metrics, degree and MLS appearances, rank power components the least critical because of their topology. Degree is based on the number of connections a component has, and power components only have a single connection to the water components they supply. Power components are not included in MLSs, so they are ranked the least critical using the number of MLS appearances. For that reason, it is suggested in this study that neither of these measures alone should be used to assess component criticality. Each of these approaches includes similar distribution components in the most critical nodes. For example, each of the metrics includes distribution node D60 in the top 5 most critical nodes. D60 is one link away from S4, with S4 being one of two of D60's connections. Therefore, it lies on many paths between other nodes and supply node, S4.

When looking at only RAW values, two water supply components and their corresponding power supply components are ranked as the four most critical nodes in the network. A power component causes a 100% probability of failure of its corresponding water supply component. The power component is only critical in the network insofar as it affects the water supply component. Therefore, there is a redundancy in the component importance rankings when both the power supply and water supply components are indicated as the most critical based on RAW alone. Combining RAW with the centrality-based metrics eliminates the redundancy while accounting for other attributes of component criticality.

For the application water system and its dependencies on power, the Degree + RAW combination metric is recommended to best assess the component criticality. This measure captures the topology of the network through degree centrality as well as the reliability of individual components through RAW.

To further assess the outcomes of using the Degree + RAW combination metric compared to the RAW metric alone, the differences in rankings between the two are shown in Figure 22. Diamonds represent the difference in ranking between using Degree + RAW and RAW. In the plot, negative values indicate that the component is more critical using the combination metric compared to the RAW metric alone.

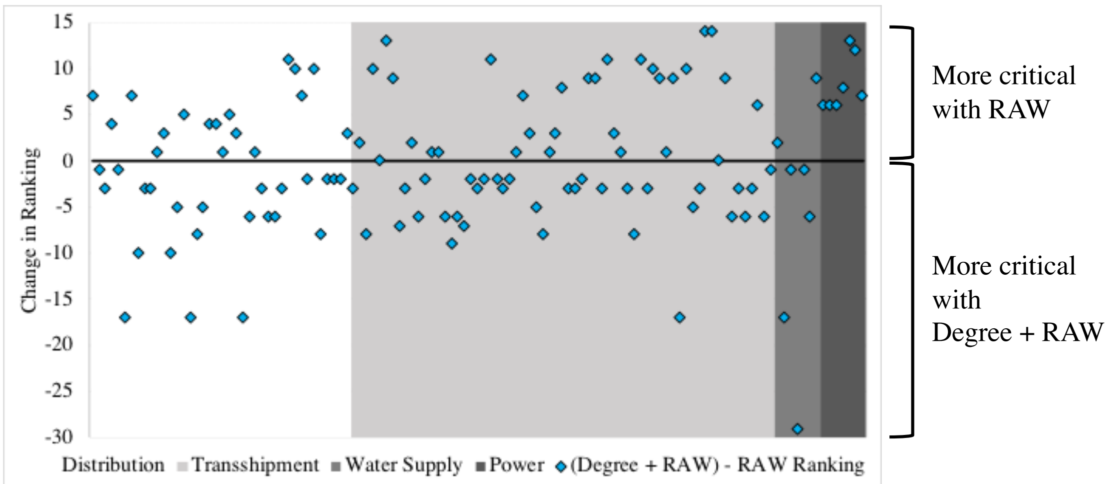


Figure 22. Comparison of rankings between Degree + RAW and RAW.

All of the power components have a positive difference between the combination metric and RAW, indicating that they are more critical using the RAW metric by itself. However, the water supply components are generally more critical using the combination metric. Water supply components are intuitively highly critical, as they are required to provide the resource that is then delivered through the network. At the same time, the RAW rankings include the unwanted redundancy between the power components and their corresponding water supply components. This highlights the importance of using a combination metric that considers additional centrality attributes when assessing component criticality.

To assess how the rankings change for components of varying characteristics, the differences in rankings between the combination Degree + RAW metric and the RAW metric alone is also compared to the component's number of connections. This is shown in Figure 23, where points represent the change in ranking of a component on the x-axis and the number of connections of the component on the y-axis.

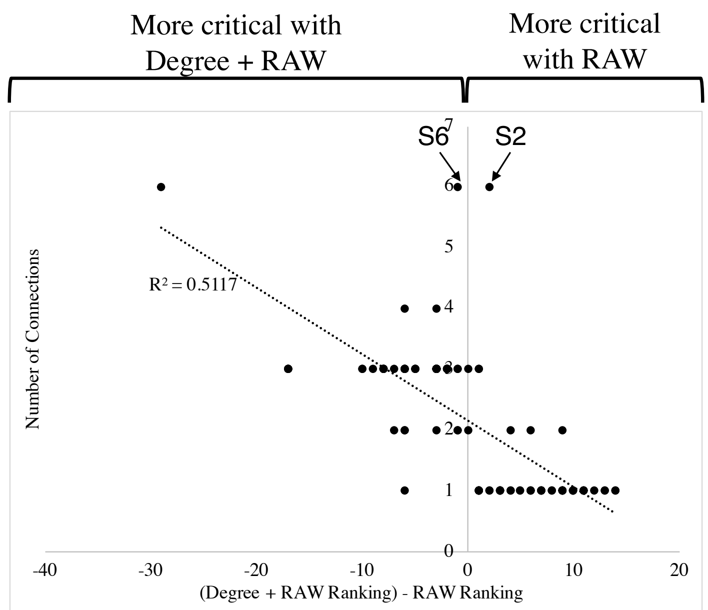


Figure 23. Comparison of change in ranking from Degree + RAW to RAW and number of connections.

This plot shows a negative relationship between the degree and change in ranking, meaning that, as expected, components with more connections increase in criticality with the combination metric. Previous studies have found centrality-based measures such as number of connections to be important indicators of component criticality. The results from using the combination metric demonstrate its ability to account for the centrality-based attributes of the components that the RAW metric alone does not capture. The R^2 value between the change in ranking and the number of connections is 0.5117. This is affected by two outliers, S2 and S6, which have many connections but little change in ranking. These are two water supply components that remain critical in both the combination metric and RAW alone rankings

6.4 Impact of Interdependencies

To quantitatively assess the impact of interdependencies on component ranking, the change in both the values of the metrics and the ranking for each component is assessed with and without consideration of dependencies of the water system on the power network. The change is calculated as the difference between the metric value or ranking with consideration of the power network and the corresponding value or ranking without consideration of power. In assessing how the value of the metric changes, for each of the metrics, a higher value indicates a larger change in the metric from including interdependencies. Components with a positive difference in the component importance measure value are more critical when the power network is included and components with a negative difference in the measure are more critical when the power network is not included. The opposite is true in assessing the impact of interdependencies on component rankings. Components with a positive difference in rank are more critical without power and components with a negative difference in rank are more critical with consideration of power.

6.4.1 Degree Centrality

For degree centrality, the change in the degree of each node with and without including dependence on power results in either zero or one. This is shown in Figure 24Figure 24. Change in degree with and without consideration of dependencies on power. The same convention for background colors used in the previous plots is used here. Diamonds represent the change in degree for each node. All of the transshipment and distribution nodes have no change in degree. The water supply nodes each have a change in degree of one when the connection to a power component is removed from the system studied. Four supply nodes (S1, S2, S4, and S7) move down the most in criticality when interdependencies are excluded from the analysis. These four water supply nodes have one to two connections compared to the five links of the other water supply nodes. S3, S5, and S6 remain the three most critical nodes with and without consideration of the power network.

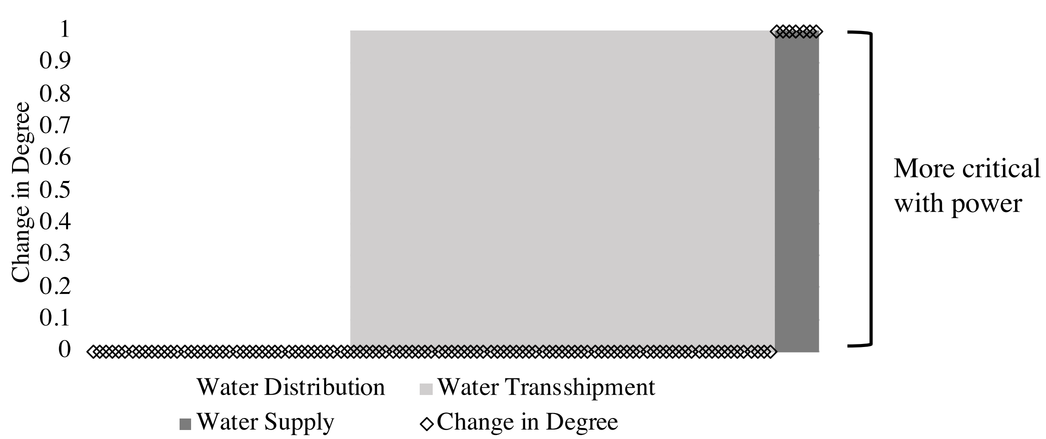


Figure 24. Change in degree with and without consideration of dependencies on power.

6.4.2 MLS Appearances

Rankings by MLS appearances are exactly the same with and without consideration of dependence on power because power components are not included in MLSs. If more information were learned about the power distribution system, a more robust network could be developed that also includes MLSs for the power system. In this case, it is likely that power components would move up in the component criticality rankings.

6.4.3 RAW

RAW values are compared with and without consideration of interdependencies as shown in Figure 25. Most components have a negative difference in the RAW value from the network with consideration of power compared to without consideration of power. I.e., most of the components in the network contribute to a larger increase in system failure probability in the case of their failure when power components are not included in the BN, compared to when power components are included in the BN.

Conversely, all of the water supply components have a positive change in RAW value, meaning that they contribute more to an increase in system failure probability in the case of their failure when power is considered as opposed to when power is not considered. The increase in RAW is due to increases in failure probabilities of water supply components given power component failures, as well as to increased updated power component failure probabilities in the interdependent case. If a water supply component has failed, its parent power component has an increased probability of failure as the BN model updates the distributions of its parent nodes. For this reason, all of the water supply components move down in criticality in the network without power dependencies with the exception of one supply component, S6, which remains at the same rank. The top ten most critical components with and without consideration of interdependencies are shown in Table 10.

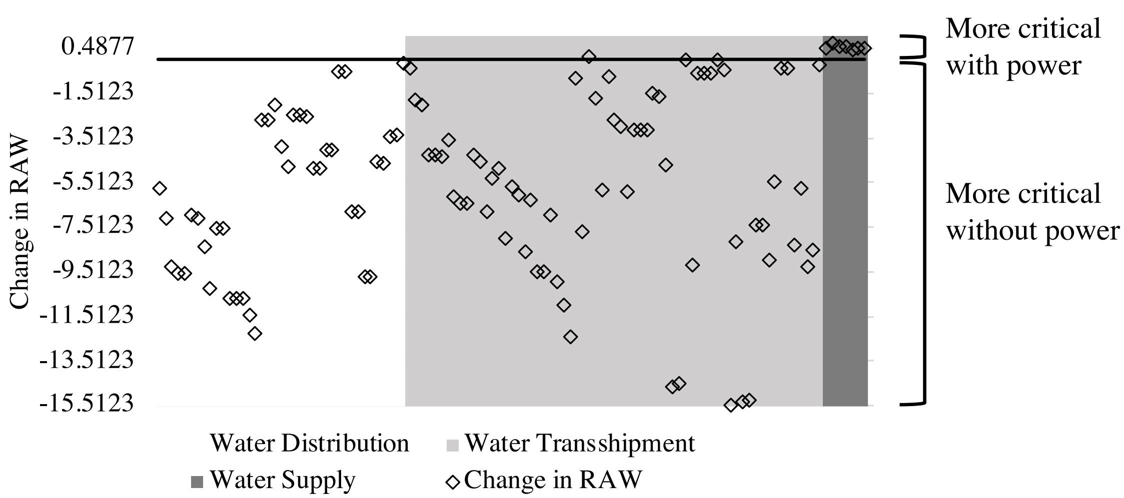


Figure 25. Change in RAW with and without consideration of dependencies on power.

Table 10. Top Ten RAW Rankings With and Without Consideration of Interdependencies.

With Interdependencies	Without Interdependencies
S6	S6
D60	D60
S4	D38
S61	D37
D63	D5
D38	D2
D37	D79
P6	D34
D65	D36

6.4.4 Degree and RAW Combination Metric

For the first combination metric using degree and RAW, most of the distribution and transshipment components have a negative difference in the measure as shown in Figure 26. I.e., most of the components contribute to a larger increase in system failure probability when they are assumed to have failed with consideration of interdependencies compared to without consideration of interdependencies. In contrast, all of the water supply components have a positive change in the degree and RAW metric. This is consistent with the RAW metric alone as well as degree alone. As supply components are critical to provide the resource that flows through the network to distribution components, this shows the importance of considering interdependencies when prioritizing repair, retrofit, or replacement actions. Table 11 shows the top ten rankings with and without considering interdependencies.

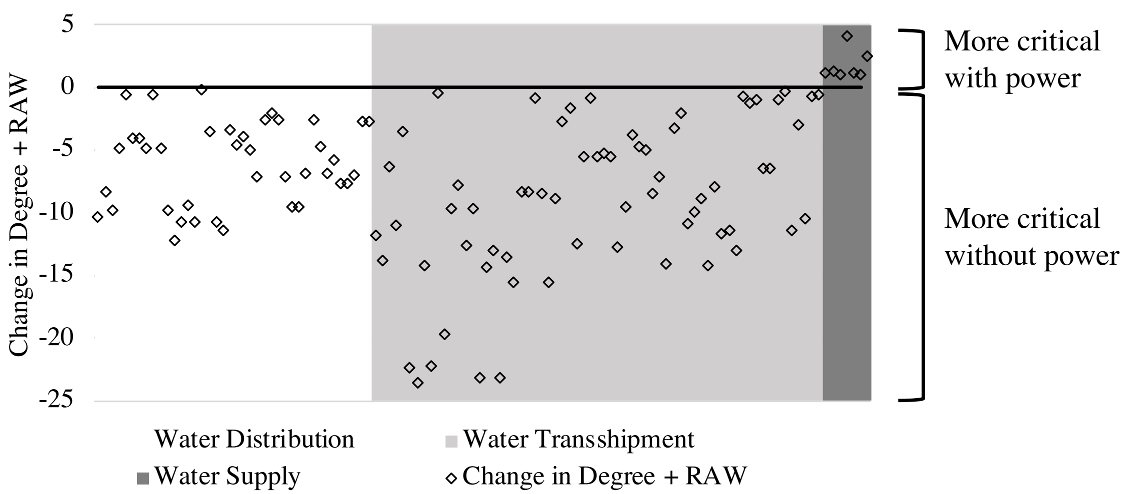


Figure 26. Change in Degree + RAW with and without consideration of dependencies on power.

Table 11. Top Ten Degree + RAW Rankings With and Without Consideration of Interdependencies.

With Interdependencies	Without Interdependencies
S6	S6
D60	D60
S4	D38
D61	D37
D63	D5
D38	D2
D37	D79
P6	D34
D65	D36
S5	D61

6.4.5 MLS Appearances and RAW Combination Metric

For the metric combining MLS appearances and RAW, 85 of the components have negative differences in the MLS and RAW metric as shown in Figure 27. Most of the water supply components have a slight positive change in the measure, with the exception of S7, which has no change. Two water supply components, S4 and S6, have the most positive change measure, indicating that they are more critical when dependence on power is considered. Again, this highlights the importance of considering interdependencies when prioritizing maintenance and repair as water supply components are critical in maintaining flow through a network. The top ten rankings with and without considering interdependencies are shown in Table 12.

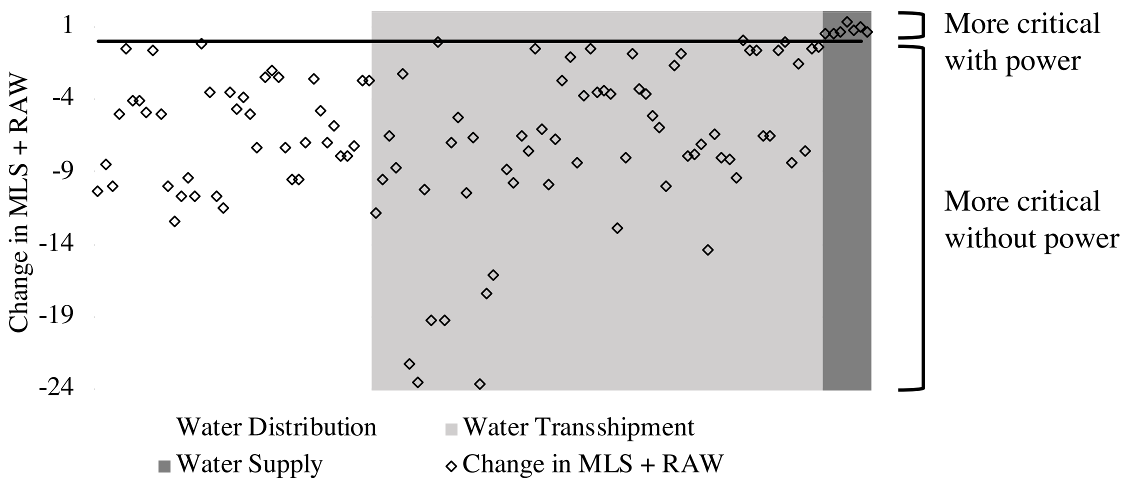


Figure 27. Change in MLS Appearances + RAW with and without consideration of dependencies on power.

Table 12. Top Ten MLS Appearances + RAW Rankings With and Without Consideration of Interdependencies.

With Interdependencies	Without Interdependencies
S6	S6
S4	S4
D60	D60
D61	D38
D38	D2
D63	D61
D37	D63
D4	D79
D65	D29
D2	D65

6.5 Component Importance Ranking for Resilience

Analyses using the framework show the effects of individual component performance on network performance taking into account the complex interdependencies that exist between infrastructure systems. Understanding the effects of the interdependencies on the fragility of the systems supports decision making in their design, management, and restoration to create more resilient critical infrastructure.

The model is useful before a hazard occurs to assess where the greatest extent of damage is possible and where to invest resources to prevent large outages. For example, infrastructure owners and decision makers can use the component criticality rankings to prioritize where the most immediate inspections should occur to determine if the most critical components are prepared for a hazard. The model is useful during a hazard to determine where to disperse resources and repair crews to bring the most customers or the most critical customers back online as quickly as possible. An emergency manager can input outages and dispatch crews and resources to the most vulnerable areas and zones with critical facilities. Finally, the model is useful after a hazard to prioritize components for interventions to prevent similar incidents and impacts from occurring again in the future. If an infrastructure operator observes cascading failures or other large outages, they can allocate resources to reinforce areas of vulnerability or create redundancies to prevent those failures in the future. Those repair and reinforcement actions can be added to the model to analyze the potential effects throughout the interdependent network.

CHAPTER 7. ALGORITHMIC PERFORMANCE

7.1 Introduction

One of the key contributions of this work is to provide computationally tractable probabilistic modeling of interdependent infrastructure systems. To demonstrate the performance of the proposed algorithms, this chapter provides a comparison of the computation time and accuracy of the framework versus prior approaches. These include MLS enumeration, BN modeling, and a comparison to Monte Carlo simulation for the same example system.

7.2 Performance Comparison to Prior Approaches

7.2.1 MLS Enumeration

To further assess performance of the proposed methodology, it is compared to that of prior approaches in several steps of the framework. The MLS enumeration takes 2.19 seconds using the proposed method. An algorithm has been developed to enumerate the complement to MLSs, MCSs by Mishra, Saifi, and Chaturvedi (2016). This is used as a comparison metric for the MLS enumeration algorithm. In the previous study, the authors propose an algorithm to identify MCSs that uses the connectivity matrix of a graph to check the connection between nodes in a network as nodes are progressively removed. The largest system that this algorithm was tested on in the study contained 21 nodes and 26 links. The enumeration of the MCSs took approximately 2,600 seconds. This is over 1,000 times longer for a network that is approximately five times smaller than the application used in this study. The MLS formulation presented allows expansion of the number of components that are included in the network with increased computational efficiency compared to other methods of identifying minimum sets in a network.

7.2.2 BN Model

Prior approaches to modeling interdependent infrastructure systems using BNs have focused on network characteristics at the global level rather than including system topologies from the component level to study system reliability and prioritize repair and retrofit for components. Therefore, inference examples in this dissertation are not comparable to works such as Aung & Watanabe (2010) and Di Giorgio & Liberati (2011). A BN approach without the MLS formulation is explored by Schaberreiter et al. (2013). However, the study is applied to a system of four infrastructure component nodes and four service nodes. The approach is not scalable to infrastructure systems of the scale used in the application.

7.2.3 BN Inference

Finally, the performance of the proposed methodology is compared to results from Monte Carlo simulation. Samples of probabilities of failure for each component were selected based on hazard occurrence using the same probabilities as described in the application for the proposed model. The probability of failure of each component given that the hazard occurs is 10^{-2} and the probability of failure given that the hazard does not occur is 10^{-4}. The failure or survival of each component was used to update the survival or failure of each MLS. These updated MLSs were then used to update the survival or failure of the nodes that depend on them. The outcome was the probability of survival of each component node. The Monte Carlo was performed using 10^3, 10^4, 10^5, 10^6, and 10^7

simulations. The calculated probabilities of survival of all components are shown in Figure 28.

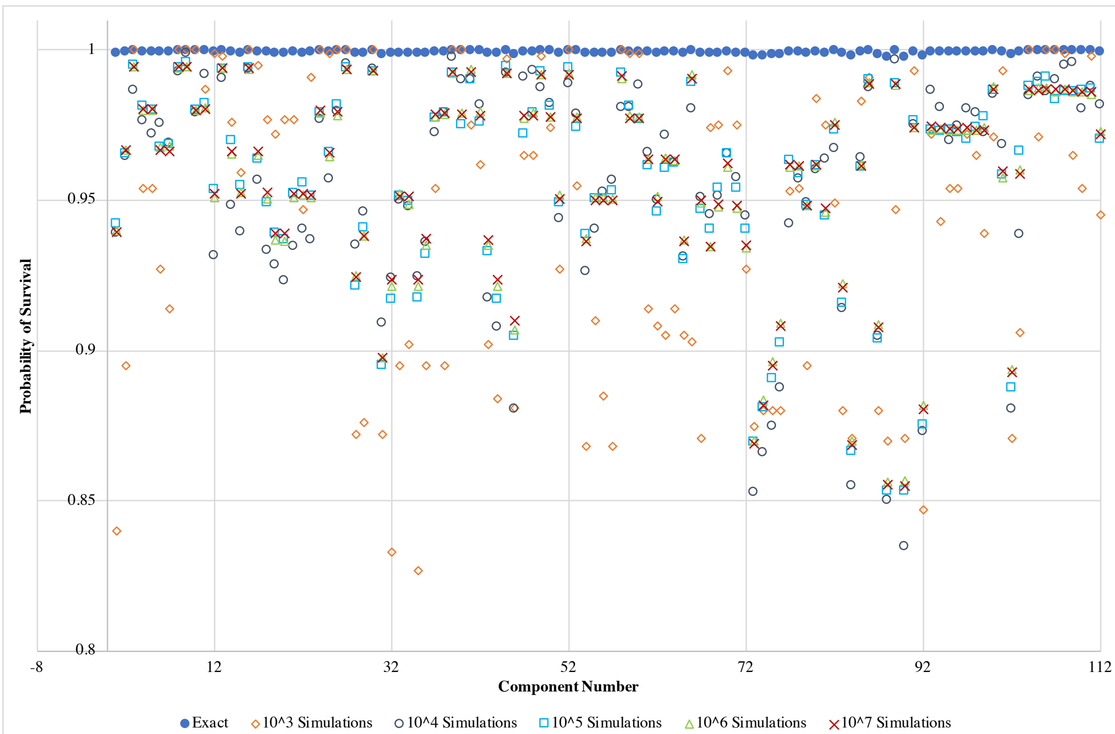


Figure 28. Comparison of results from proposed methodology to Monte Carlo simulations.

The solid circles represent the probabilities of survival of each component calculated using the proposed framework. These are the exact solutions. The open diamonds, circles, squares, and triangles represent the probabilities obtained from 10^3, 10^4, 10^5, and 10^6 simulations, respectively, with ×s representing results from 10^7 simulations. To compare the error in Monte Carlo Simulation, two smaller conditional probabilities of failure were also assessed as shown in Table 13. The second set of failure probabilities is 10^{-4} given a hazard and 10^{-6} given no hazard. The third set of failure probabilities is 10^{-6} given a hazard and 10^{-8} given no hazard. The computation times required for each method and the average percent errors over all components are shown in Table 13.

Table 13. Comparison of performance of proposed methodology to Monte Carlo simulations.

	Proposed	**10^3 Simulations**	**10^4 Simulations**	**10^5 Simulations**	**10^6 Simulations**	**10^{-7} Simulations**
Time (seconds)	88.66	0.17	0.67	5.31	55.49	3684.67
Average Error 10^{-2},10^{-4}	0%	5.49%	4.36%	4.18%	4.14%	4.13%
Average Error 10^{-4},10^{-6}	0%	100%	100%	82.04%	37.36%	23.93%
Average Error 10^{-6},10^{-8}	0%	100%	100%	98.58%	81.59%	79.40%

As expected for Monte Carlo, the error decreases with an increase in the number of simulations. However, the average errors decrease slowly for the probabilities of failure of 10^{-2} and 10^{-4} as the orders of magnitude of the numbers of simulations increase. The computation time for 10^7 simulations was much higher than the proposed approach and also required a large amount of memory, a 112×10^7 matrix representing the state of each component for every simulation. Results from the simulation approach do not capture the small differences in failure probabilities that the exact solutions obtained using the proposed method give. The errors are significant given the importance of detailed granularity in the probabilities of survival in the results, particularly if they are used to rank component criticality in the interdependent network. When small probabilities of failure are assessed, such as 10^{-4} and 10^{-6}, errors increase in magnitude. These probabilities of failure are reasonable for many infrastructure components that are not experiencing increased demands given a hazard.

7.3 Summary of Performance Measures

Table 14 shows a summary of the time taken to compute each of the above steps of the framework as well as the amount of memory required. Through participating in the National Science Foundation's Innovation-Corps program, time taken and memory requirements were found to be useful performance metrics for infrastructure owners. Additional measures such as amount of money saved through implementing actions recommended by the analyses would also be useful. Through speaking to over 100 people in the utility industry, the addition of consideration of external dependencies would be beneficial in their existing methods to determine which components to invest in for repair, maintenance, and addition of redundant paths.

Table 14. Summary of time and memory requirements.

	Input	**MLS Enumeration**	**BN Construction**	**BN Inference**
Time (seconds)	------	2.19	88.66	4.00
Memory	1.3 MB	2 KB	995 KB	---------------

CHAPTER 8. FUTURE WORK

8.1 Introduction

This chapter provides an outline of future work that can be expanded upon from this work including application of the framework to capacitated networks, development of quantified resilience metrics, and application of the framework to macro-scale infrastructure systems.

8.2 Application to Capacitated Networks

The MLS formulation described in this dissertation is connectivity-based and does not account for flow through or capacity of a system. Rather, the focus is on connectivity of the nodes in the network. Other MLS formulations, such as the max-flow min-cut theorem, can be explored to capture the flow between supply and distribution nodes and allow for more detailed analysis of capacitated networks.

Incorporation of directionality and capacity in the network would require updates to the definition of CPTs in the framework. Rather than representing components as binary (i.e., either surviving or failing), components would be represented as continuous, where states would represent the level of flow throughout the node. The CPT for each component would represent the probability of maintaining a flow level given the states of parent nodes, such as occurrence of a hazard or survival of a given number of MLS parents.

8.3 Quantification of Interdependent Infrastructure Resilience

Several prior studies have proposed methods to quantify the resilience of single infrastructure systems. A review of these proposed methods is provided in Hosseini, Barker, and Ramirez-Marquez (2016). However, these measures do not account for interdependencies between infrastructure systems. This model can be expanded upon to measure the resilience of infrastructure systems while accounting for the complex interdependencies between the infrastructure systems.

One option would be to simulate a hazard on the system and measure the proportion of components that fail in each service area. This would account for the preparation aspect of resilience. In response and recovery, the component importance rankings could be used to find a prioritized order of repair. A time dimension could be added to assess the results of specific recovery actions over time using a dynamic BN. Additionally, cost nodes could be added to account for the cost of repairing each node, in order to help optimize recovery actions.

8.4 Application to Larger-Scale Infrastructure Systems

The application of the framework to larger infrastructure systems can be assessed , including through investigations using the existing example network. The City of Atlanta provided data including pipes of one inch and above. The application in this study focused on pipes greater than or equal to 18 inches. To test the performance of the algorithm on a denser system, smaller pipes would be added for analysis. To assess the performance of the algorithm, one would assess the time and memory requirements of increasing numbers of MLSs with potentially more components in each MLS, a larger interdependent adjacency matrix, definition of additional CPTs, and inference on the larger BN.

As described previously, representative sub-BNs can be used to expand the BN to larger-scale systems. In the application provided in this study, for example, some nodes are compressed. To account for every possible node, sub-networks can be used to represent smaller groups of nodes. These sub-groups would then be connected to the network as a whole, so the entire network is still represented. Future work can also address how best to form these sub-groups of nodes.

CHAPTER 9. SUMMARY AND CONCLUSIONS

9.1 Introduction

This chapter provides a summary of the specific contributions of the study.

9.2 Summary of Contributions

This study describes the development and validation of a new framework to perform probabilistic vulnerability analyses of interdependent infrastructures. The proposed approach is presented, including algorithms to construct a BN model of the interdependent infrastructure systems. The method results in computationally efficient modeling and analysis of large infrastructure networks with exact inference possible over any number of system states. Inference over the network enables scenario-based analyses and prioritization from the component level for repair, replacement, or reinforcement decisions. This is useful to increase system resilience before a hazard occurs to assess where the greatest extent of damage is possible and where to invest resources to prevent large outages. The model is useful during a hazard to determine where to disperse resources and repair crews to bring the most customers or the most critical customers back online as quickly as possible. Finally, the model is useful after a hazard to prioritize components for interventions to prevent similar incidents and impacts from occurring again in the future.

The BN formulation accounts for uncertainty within the system as well as the interdependencies between different infrastructure systems. In the example, dependence on power and geographic interdependencies are represented. More information on the connectivity and locations of components in the power network would allow for expansion to two-way interdependencies. Uncertainties in both individual component failure probabilities and the probabilistic connections between components are included in the model. Compared to previous input-output-based methods (Leontief, 1951; Rose & Miernyk, 1989) requiring a large amount of data, only simple inputs of basic component characteristics of location, type, connectivity, and initial failure probabilities are required. Dimensionality reduction algorithms, including for minimum link set, super-component, and cycle identification, allow the model to include hundreds of component nodes while remaining computationally efficient without making any approximating assumptions. The proposed modeling approach and framework for analysis enables us to create more reliable and resilient networks by understanding where vulnerabilities in the system exist and the areas where investing resources will lead to the greatest improvements in predicted system outcomes.

The model created is computationally tractable and scalable. Where prior work, such as Bobbio et al. (2001) and Kim (2011) can only be applied to small networks, the proposed work can be applied to much larger systems and sub-networks can be used to expand the model even further. I.e., sub-networks can be built to represent some portion of a larger network, such as part of the American electric grid, and then sub-networks can be connected to create a representative model on a macro-scale. Previous work has used BNs to model interdependent infrastructures at a global scale, such as Schaberreiter et al. (2013) where high-level system metrics are nodes in the BN and Aung & Watanabe (2010) where entire infrastructure sectors are nodes in the BN. In the proposed framework, systems are modeled from the component scale. This allows the importance of individual components to be assessed as opposed to prior work focusing only on identifying the most critical infrastructure type. With the proposed modeling and assessment of interdependent infrastructure systems, the objective is to identify the most critical infrastructure components with consideration of interdependencies between systems.

The impacts of interdependencies on analyses of component criticality are assessed using centrality-based and reliability-based component importance measures for interdependent infrastructure systems. Two centrality-based metrics, degree centrality and MLS appearances, and one reliability-based metric, RAW, are applied to an example water distribution system in Atlanta, Georgia, and its dependencies on power. Two combination metrics are proposed – Degree + RAW and MLS Appearances + RAW – to capture the importance of network topology on component importance as well as the reliability of the components themselves. Metrics are assessed for the system, including where water and power supply components lie in the rankings as well as other attributes of the critical components. Each of the rankings places some or most water supply components at or near the highest criticality. For the centrality-based metrics, power supply components, which affect the water supply nodes, are among the least critical in the rankings. This highlights the importance of considering individual component reliabilities in determining component criticality. The water supply components have higher criticality when dependence on power is considered compared to when power is not considered. Including the dependence results in identification of the importance of water supply components in the network.

When RAW is considered by itself, water supply components and their corresponding power supply components appear side-by-side in the rankings. This can be redundant, as the power component's outage will cause 100% probability of failure of the water supply component. The combination metrics are then useful to eliminate that redundancy, while still accounting for the importance of the power components. The most appropriate metric in this application is the metric combining degree centrality and RAW. Because of its wide applicability if more information were known about the power network, the combination metric with MLS appearances and RAW would also be a beneficial metric for component importance ranking. These combination metrics build upon previous applications of centrality-based metrics on their own (e.g., Comfort & Haase, 2006; Stergiopoulos et al, 2015) and reliability-based metrics on their own (e.g., Oliveira, Mota De Sá, & Ferreira, 2014) to account for individual component reliability and give higher weight to the topology of the network. From the results, the combination metric accounts for nodes with a higher degree more than RAW alone, showing that the topology is explicitly considered when assessing component criticality.

The rankings using each metric are compared when system interdependencies are included versus when they are not. Of the five metrics, four have different components in the top ten most critical nodes. This implies that when looking at the system as a whole with its exterior dependencies, new critical components may be revealed that are not highlighted when a component is considered only in its own system. Using many of the metrics in the application, water supply components typically increase in importance with consideration of dependence on power. Many transshipment components show greater changes in the value of the component importance measure when dependencies on power are considered. These components are generally in close proximity to supply components, and therefore lead to increases in updated failure probabilities of their surrounding supply components as well.

Infrastructure systems are inherently interdependent, but often considered on their own when maintenance decisions are made. Considerations of external dependencies are necessary in making decisions to repair, replace, and reinforce component to increase overall system performance and resilience. The methodology provided in this study enables the modeling of infrastructure systems and their complex interdependencies. It allows infrastructure decision makers to understand where vulnerabilities exist in their systems and prioritize investment in order to create stronger, more resilient systems.

REFERENCES

Akhtar, R., & Santos, J. R. (2013). Risk-based input-output analysis of hurricane impacts on interdependent regional workforce systems. *Natural Hazards, 65*(1), 391-405.

Albert, R., Albert, I., & Nakarado, G. L. (2004). Structural vulnerability of the North American power grid. *Physical Review E, 69*, 025103.

American Society of Civil Engineers. (2017). *2017 Report Card for America's Infrastructure*. Retrieved from http://www. infrastructurereportcard.com

Andersson, G., Donalek, P., Farmer, R., Hatziargyriou, N., Kamwa, I., Kundur, P., Martins, N., Paserba, J., Pourbeik, P., Sanchez-Gasca, J., Schulz, R., Stankovic, A., Taylor, C., & Vittal, V. Causes of the 2003 Major Grid Blackouts in North America and Europe, and Recommended Means to Improve System Dynamic Performance. *IEEE Transactions on Power Systems, 20*(4), 1922-1928.

Applegate, C. J., & Tien, I. (2018). Framework for probabilistic vulnerability analysis of interdependent infrastructure systems. *Journal of Computing in Civil Engineering*.

Applegate, C. J., & Tien, I. (Submitted in July 2018). Impacts of infrastructure interdependencies on assessments of component criticality through combined centrality- and reliability-based measures. *Reliability Engineering and System Safety*.

Aung, Z. Z., & Wantanabe, K. (2010). Modeling Inoperability Propagation Using Bayesian Networks. In M. T., S. Shenoi, & (eds), *Critical Infrastructure Protection IV. ICCIP 2010. IFIP Advances in Information and Communication Technology* (Vol. 342, pp. 199-212). Springer, Berlin, Heidelberg.

Barker, K., Ramirez-Marquez, J. E., & Rocco, C. M. (2013). Resilience-based network component importance measures. *Reliability Engineering & System Safety, 117*, 89-97.

Baroud, H., & Barker, K. (2018). A Bayesian kernel approach to modeling resilience-based network component importance. *Reliability Engineering and System Safety, 170*, 10-19.

Baroud, H., Barker, K., & Ramirez-Marquez, J. (2013). Measures of inland waterway network resilience. *INCOSE International Symposium*, *23 (1)*, pp. 1354-1367.

Bensi, M., Der Kiureghian, A., & Straub, D. (2013). Efficient Bayesian network modeling of systems. *Reliability Engineering and System Safety, 112*, 200-213.

Bobbio, A., Portinale, L., Minichino, M., & Ciancamerla, E. (2001). Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliability Engineering and System Safety, 71*(3), 249-260.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., & von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. *EERI Spectra Journal, 19*(4), 733-752.

Cadini, F., Zio, E., & Petrescu, C. A. (2008, October). Using centrality measures to rank the importance of the components of a complex network infrastructure. In R. Setola, & S. Geretshuber, *International Workshop on Critical Information Infrastructures Security* (pp. 155-167). Rome, Italy: Springer-Verlag Berlin Heidelberg.

Castillo, E., Grande, Z., & Calviño, A. (2016). Bayesian Networks-Based Probabilistic Safety Analysis for Railway Lines. *Computer-Aided Civil and Infrastructure Engineering, 31*(9), 681-700.

Cheng, M.-Y., & Hoang, N.-D. (2014). Slope collapse prediction using Bayesian framework with k-nearest neighbor density estimation: case study in Taiwan. *Journal of Computing in Civil Engineering, 30*(1), 04014166.

Chou, C.-C., & Tseng, S.-M. (2010). Collection and analysis of critical infrastructure interdependency relationships. *Journal of Computing in Civil Engineering, 24*(6), 539-547.

Codetta-Raiteri, D., Bobbio, A., Montani, S., & Portinale, L. (2012). A dynamic Bayesian network based framework to evaluate cascading effects in a power grid. *Engineering Applications of Artificial Intelligence, 25*(4), 683-697.

Comfort, L. K., & Haase, T. W. (2006). Communication, coherence, and collective action: The impact of Hurricane Katrina on communications infrastructure. *Public Works Management & Policy, 10*(4), 328-343.

Crucitti, P., Latora, V., & Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E, 73*(3), 036125.

Der Kiureghian, A., & Song, J. (2008). Multi-scale reliability analysis and updating of complex systems by use of linear programming. *Reliability Engineering & System Safety, 93*(2), 288-297.

Di Giorgio, A., & Liberati, F. (2011, June). Interdependency modeling and analysis of critical infrastructures based on Dynamic Bayesian Networks. *Control & Automation (MED), 2011 19th Mediterranean Conference* (pp. 791-797). Aquis Corfu Holiday Palace, Corfu, Greece: IEEE.

Dueñas-Osorio, L., Craig, J., & Goodno, B. J. (2007). Seismic response of critical interdependent networks. *Earthquake Engineering and Structural Dynamics, 36*(2), 285-306.

Dunn, S., & Wilkinson, S. M. (2012). Identifying Critical Components in Infrastructure Networks Using Network Topology. *Journal of Infrastructure Systems, 19*(2), 157-165.

Espiritu, J. F., Coit, D. W., & Prakash, U. (2007). Component criticality importance measures for the power industry. *Electric Power Systems Research, 77*(5-6), 407-420.

Franchin, P., & Cavalieri, F. (2015). Probabilistic assessment of civil infrastructure resilience to earthquakes. *Computer-Aided Civil and Infrastructure Engineering, 30*(7), 583-600.

González, A. D., Dueñas-Osorio, L., Sánchez-Silva, M., & Medaglia, A. L. (2016). The interdependent network design problem for optimal infrastructure system restoration. *Computer-Aided Civil and Infrastructure Engineering, 31*(5), 334-350.

Grande, Z., Castillo, E., Mora, E., & Lo, H. K. (2017). Highway and Road Probabilistic Safety Assessment Based on Bayesian Network Models. *Computer-Aided Civil and Infrastructure Engineering, 32*(5), 379-396.

Grubesic, T. H., Matisziw, T. C., Murray, A. T., & Snediker, D. (2008). Comparative approaches for assessing network vulnerability. *International Regional Science Review, 31*(1), 88-112.

Guidotti, R., Chmielewski, H., Unnikrishnan, V., Gardoni, P., McAllister, T. P., & van de Lindt, J. (2016). Modeling the resilience of critical infrastructure: The role of network dependencies. *Sustainable and Resilient Infrastructure, 1*(3-4), 153-168.

Halfawy, M. R. (2008). Integration of municipal infrastructure asset management processes: challenges and solutions. *Journal of Computing in Civil Engineering, 22*(3), 216-229.

Hernandez-Fajardo, I., & Dueñas-Osorio, L. (2013). Probabilistic study of cascading failures in complex interdependent lifeline systems. *Reliability Engineering and System Safety, 111*, 260-272.

Hosseini, S., & Barker, K. (2016). Modeling infrastructure resilience using Bayesian networks: A case study of inland waterway ports. *Computers & Industrial Engineering, 93*, 252-266.

Hosseini, S., Barker, K., & Ramirez-Marquez, J. E. (2016). A review of definitions and measures of system resilience. *Reliability Engineering and System Safety, 145*, 47-61.

Jiang, X., Bai, R., Atkin, J., & Kendall, G. (2017). A scheme for determining vehicle routes based on Arc-based service network design. *INFOR: Information Systems and Operational Research, 55*(1), 16-37.

Johansen, C., & Tien, I. (2018). Probabilistic multi-scale modeling of interdependencies between critical infrastructure systems for resilience. *Sustainable and Resilient Infrastructure, 3*(1), 1-15.

Johansen, C., Horney, J., & Tien, I. (2016). Metrics for evaluating and improving community resilience. *ASCE Journal of Infrastructure Systems, 23*(2), 04016032.

Johansson, J., & Hassel, H. (2010). An approach for modelling interdependent infrastructures in the context of vulnerability analysis. Reliability Engineering and System Safety 95, 1335-1344.

Khakzad, N. (2015). Application of dynamic Bayesian network to risk analysis of domino effects in chemical infrastructures. *Reliability Engineering & System Safety, 138*, 263-272.

Kim, M. C. (2011). Reliability block diagram with general gates and its application to system reliability analysis. *Annals of Nuclear Energy, 38*(11), 2456-2461.

Leontief, W. W. (1951). Input-output economics. *Scientific American, 185*(4), 15-21.

Leu, S.-S., & Bui, Q.-N. (2016). Leak prediction model for water distribution networks created using a Bayesian network learning approach. *Water Resource Management, 30*(8), 2719-2733.

Li, J., & He, J. (2002). A recursive decomposition algorithm for network seismic reliability evaluation. *Earthquake Engineering & Structural Dynamics, 31*(8). 1525-1539.

Lim, H.-W., Song, J., Kurtz, N. (2015). *Earthquake Engineering & Structural Dynamics, 44*, 355-369.

Mahadevan, S., Zhang, R., & Smith, N. (2001). Bayesian Networks for system reliability reassessment. *Structural Safety, 23*(3), 231-251.

Mishra, R., Saifi, M. A., & Chaturvedi, S. K. (2016). Enumeration of minimal cutsets for directed networks with comparative reliability study for paths or cuts. *Quality and Reliability Engineering International, 32*(2), 555-565.

Murphy, K. P. (2001). The Bayes Net Toolbox for Matlab. *Computing Science and Statistics, 33*(2), 1024-1034.

Naidu, A. S., Soh, C. K., & Pagalthivarthi, K. V. (2006). Bayesian network for E/M impedance-based damage identification. *Journal of Computing in Civil Engineering, 20*(4), 227-236.

Oliveira, C. S., Mota De Sá, F., & Ferreira, M. A. (2014, August). Disaster Prevention Strategies Based on the Level of Risk. (pp. 25-29). Istanbul: Proceedings 2nd European conference on earthquake engineering and seismology.

Omer, M., Mostashari, A., & Lindemann, U. (2014). Resilience Analysis of Soft Infrastructure Systems. *Procedia Computer Science, 28*, 565-574.

Ouyang, M. (2014). Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability Engineering & System Safety, 121*, 43-60.

Patterson, S. A., & Apostolakis, G. E. (2007). Identification of critical locations across multiple infrastructures for terrorist actions. *Reliability Engineering & System Safety, 92*(9), 1183-1203.

Pinnaka, S., Yarlagadda, R., & Çetinkaya, E. K. (2015, March). Modelling robustness of critical infrastructure networks. *In Design of Reliable Communication Networks (DRCN), 2015 11th International Conference* (pp. 95-98). IEEE.

PlaNYC. (2013). A Stronger, More Resilient New York. New York, NY.

President′s Commission on Critical Infrastructure Protection (PCCIP). (1997). *Critical foundations: protecting America's infrastructures: the report of the President 's commission on critical infrastructure protection. Rep. No. 040-000-00699- 1.* Washington, D.C.: U.S. Government Printing Office.

Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine, 21*(6), 11-25.

Rose, A., & Miernyk, W. (1989). Input-output analysis: The first fifty years. *Economic Systems Research, 1*(2), 229-272.

Santos, J. R., Pagsuyoin, S. T., Herrera, L. C., Tan, R. R., & Yu, K. D. (2014). Analysis of drought risk management strategies using dynamic inoperability input–output modeling and event tree analysis. *Environmental Systems and Decisions, 34*(4), 492-506.

Satumtira, G., & Dueñas-Osorio, L. (2010). Synthesis of Modeling and Simulation Methods on Critical Infrastructure Interdependencies Research. In K. Gopalakrishnan, & S. Peeta, *Synthesis of Modeling and Simulation Methods on Critical Infrastructure Interdependencies Research* (pp. 1-51). Springer-Verlag Berlin Heidelberg.

Schaberreiter, T., Bouvry, P., Röning, J., & Khadraoui, D. (2013). A Bayesian network based critical infrastructure risk model. *EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation II* (pp. 207-218). Springer, Berlin, Heidelberg.

Shin, Y. Y., & Koh, J. S. (1998). An algorithm for generating minimal cutsets of undirected graphs. *Korean Journal of Computing & Applied Mathematics, 5*(3), 681-693.

Stergiopoulos, G., Kotzanikolaou, P., Theocharidou, M., & Gritzalis, D. (2015). Risk mitigation strategies for Critical Infrastructures based on graph centrality analysis. *International Journal of Critical Infrastructure Protection, 10*, 34-44.

Tien, I. (2017). Bayesian network methods for modeling and reliability assessment of infrastructure systems. In *Risk and Reliability Analysis: Theory and Applications* (pp. 517-452). Springer International Publishing.

Tien, I., & Der Kiureghian, A. (2017). Reliability Assessment of Critical Infrastructure Using Bayesian Networks. *Journal of Infrastructure Systems, 23*(4), 04017025.

Tong, Y., & Tien, I. (2017). Algorithms for Bayesian Network Modeling, Inference, and Reliability Assessment for Multistate Flow Networks. *Journal of Computing in Civil Engineering, 31*(5), 04017051.

U.S. Department of Energy. (2014). *LARGE POWER TRANSFORMERS AND THE U.S.ELECTRIC GRID.* Infrastructure Security and Energy Restoration Office of Electricity Delivery and Energy Reliability.

UNIDSR, WMO. (2012). UN System Task Team on the Post-2015 UN Development Agenda.

United States - Computer Emergency Readiness Team. (2016). Alert (TA16-250A) The Increasing Threat to Network Infrastructure Devices and Recommended Mitigations . United States Department of Homeland Security.

Vesely, W. E., & Davis, T. C. (1985). Two Measures of Risk Importance and Their Application. *Nuclear Technology, 68*(2), 226-234.

Volkanovski, A., Čepin, M., & Mavko, B. (2009). Application of the fault tree analysis for assessment of power system reliability. *Reliability Engineering & System Safety, 94*(6), 1116-1127.

White House. (2013). Presidential Policy Directive -- Critical Infrastructure Security and Resilience.

Zhang, X., Miller-Hooks, E., & Denny, K. (2015). Assessing the role of network topology in transportation network resilience. *Journal of Transport Geography, 46*, 35-45.